# Topology

Course Notes — Harvard University — Math 131

Fall 2013

C. McMullen

## Contents



## 1 Introduction

Topology is simply geometry rendered flexible. In geometry and analysis, we have the notion of a metric space, with distances specified between points. But if we wish, for example, to classify surfaces or knots, we want to think of the objects as rubbery.

**Examples.** For a topologist, all triangles are the same, and they are all the same as a circle. For a two–dimensional example, picture a torus with a hole in it as a surface in $\mathbb{R}^3$. This space can be almost completely flattened out. Indeed, it is topologically the same as a rotary with an underpass connecting its inner and outer edges.

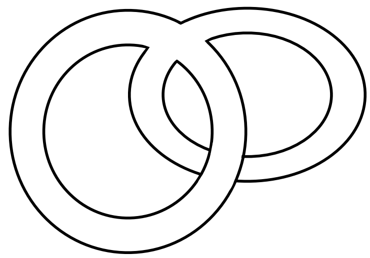


Figure 1. Rotary with underpass.

If we take an ordinary band of paper and put in a twist, it becomes a Möbius band, which seems clearly different, but how so? What happens if you cut a Möbius band down the middle? What happens if you do it again? Is a configuration of two linked circles in space fundamentally different from two unlinked circles?

**The idea of a topological space.** The property we want to maintain in a topological space is that of *nearness*. We will allow shapes to be changed, but without tearing them. This will be codified by open sets.

Topology underlies all of analysis, and especially certain large spaces such as the dual of $L^\infty(\mathbb{Z})$ lead to topologies that cannot be described by metrics.

Topological spaces form the broadest regime in which the notion of a *continuous function* makes sense. We can then formulate classical and basic theorems about continuous functions in a much broader framework.

For example, an important theorem in optimization is that any continuous function $f : [a, b] \to \mathbb{R}$ achieves its minimum at least one point $x \in [a, b]$. This property turns out to depend only on *compactness* of the interval, and not, for example, on the fact that the interval is finite–dimensional.

**Invariants.** A second agenda in topology is the development of tools to tell topological spaces apart. How is the Möbius band to be distinguished from the cylinder, or the trefoil not from the figure–eight knot, or indeed how is $\mathbb{R}^3$ different from $\mathbb{R}^4$? Our introduction to the tools of *algebraic topology* provides one approach to answer these questions.

**This course.** This course correspondingly has two parts. Part I is *point–set topology*, which is concerned with the more analytical and aspects of the theory. Part II is an introduction to *algebraic topology*, which associates algebraic structures such as groups to topological spaces.

We will follow Munkres for the whole course, with some occassional added topics or different perspectives.

We will consider topological spaces *axiomatically.* That is, a topological space will be a set $X$ with some additional structure. Because of the generality of this theory, it is useful to start out with a discussion of set theory itself.

**Remark on writing proofs.** When you hit a home run, you just have to step once on the center of each base as you round the field. You don't have to circle first base and raise a cloud of dust so the umpire can't quite see if you touched the base but will probably give you the benefit of the doubt.

# 2 Background in set theory

**The axioms of set theory.**

- Axiom I. (Extension) A set is determined by its elements. That is, if $x \in A \implies x \in B$ and vice-versa, then $A = B$.

- Axiom II. (Specification) If $A$ is a set then $\{x \in A : P(x)\}$ is also a set.

- Axiom III. (Pairs) If $A$ and $B$ are sets then so is $\{A, B\}$. From this axiom and $\emptyset = 0$, we can now form $\{0, 0\} = \{0\}$, which we call 1; and we can form $\{0, 1\}$, which we call 2; but we cannot yet form $\{0, 1, 2\}$.

- Axiom IV. (Unions) If $A$ is a set, then $\bigcup A = \{x : \exists B, B \in A \ \& \ x \in B\}$ is also a set. From this axiom and that of pairs we can form $\bigcup\{A, B\} = A \cup B$. Thus we can define $x^+ = x + 1 = x \cup \{x\}$, and form, for example, $7 = \{0, 1, 2, 3, 4, 5, 6\}$.

- Axiom V. (Powers) If $A$ is a set, then $\mathcal{P}(A) = \{B : B \subset A\}$ is also a set.

- Axiom VI. (Infinity) There exists a set $A$ such that $0 \in A$ and $x+1 \in A$ whenever $x \in A$. The smallest such set is unique, and we call it $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$.
- Axiom VII (The Axiom of Choice): For any set $A$ there is a function $c : \mathcal{P}(A) - \{\emptyset\} \to A$, such that $c(B) \in B$ for all $B \subset A$.

**Discussion of the Axioms.** **Axiom I. (Extension).** To have a well–defined domain of discourse, the elements of sets are *also sets*.

There is a subtle point in Axiom I: what does the conclusion, $A = B$, mean anyway? In fact the idea of equality is a notion in logic rather than set theory. It means that for any logical sentence $P(x)$, $P(A)$ has the same answer as $P(B)$. For example, if $A = B$, and $A \in Y$, then $B \in Y$.

**Axiom II. (Specification).** Examples: $A \cap B = \{x \in A : x \in B\}$. $A - B = \{x \in A : x \notin B\}$. $\mathbb{R} - \mathbb{Q} =$ irrationals; $\mathbb{Q} - \mathbb{R} = \emptyset$.

$$\begin{aligned}
&\{x \in \mathbb{Z} : \exists y \in \mathbb{Z}, y + y = x\} = \text{even numbers.} \\
&\{x \in \mathbb{Z} : x/n \in \mathbb{Z}\ \forall n > 0\} = \{0\}. \\
&\{x \in \mathbb{Z} : x^2 < 0\} = \emptyset.
\end{aligned}$$

For more advanced set theory, one uses the Axiom of Replacement instead of Specification; this permits the construction of cardinals such as $\aleph_\omega$, but it is not required for most ‘mainstream’ mathematics.

Assuming at least one set $A$ exists, we can now form

$$0 = \emptyset = \{x \in A : x \neq x\},$$

but nothing else for sure. (E.g. $A$ might be $\emptyset$.)

**The Barber of Seville; Russell’s paradox.** If $X = \{A : A \notin A\}$, is $X \in X$? There is no universe: given a set $A$, set $X = \{B \in A : B \notin B\}$. We claim $X \notin A$. Indeed, if $X \in A$, then $X \in X$ iff $X \notin X$.

One solution to the classic paradox — who shaves the barber of Seville? — is of course that the barber is a woman. In the Gödel-Bernays theory, you are allowed to form $X$, but $X$ is not a set; it is called a class.

**Axiom III. (Pairs).** From this axiom and $\emptyset = 0$, we can now form $\{0, 0\} = \{0\}$, which we call 1; and we can form $\{0, 1\}$, which we call 2; but we cannot yet form $\{0, 1, 2\}$.

**Axiom IV. (Unions).** From this axiom and that of pairs we can form $\bigcup\{A, B\} = A \cup B$. Thus we can define $x^+ = x + 1 = x \cup \{x\}$, and form, for example, $7 = \{0, 1, 2, 3, 4, 5, 6\}$.

**Intersections.** If $A \neq \emptyset$, we can define $\bigcap A = \{x \, : \, \forall B \in A, x \in B\}$. Since $A$ has at least one element $B_0$, we have $\bigcap A \subset B_0$ and thus the intersection is a set. Note: $\bigcap \emptyset$ is undefined!

Examples: $\bigcap\{A\} = A$, $\bigcap\{A, B\} = A \cap B$.

**Axiom V. (Powers)** Examples: $X = \{B \in \mathcal{P}(52) : B \text{ has exactly 5 elements}\}$ is the number of possible poker hands. $|X| = 2,598,960$.

Pascal's triangle. The subsets with $k + 1$ elements of $\{1, \ldots, n\}$ can be partitioned into those that include $n$ and those that do not. Thus $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$.

**Axiom VI. (Infinity)** . We have now built up the natural numbers via set theory, and can proceed to the real numbers, functions on them, etc., with everything resting on the empty set.

Another standard assumption we have not listed is the Axiom of Extension, which asserts there is no decreasing sequence $\ldots x_3 \in x_2 \in x_1$. This implies *all* sets rest on the empty set, and we never have the *infinite loop* $x \in x$.

A kindly mathematician uncle asks his niece, "What's the highest number you know?" The niece replies, "168,000,000". The uncle asks, "But what about 168,000,001"? And the niece replies, "I was close, wasn't I?" (Hubbard)

**Moving along.** We can now define *ordered pairs* by

$$(a, b) = \{\{a\}, \{a, b\}\}.$$

Then $(a, b) = (a', b')$ iff $a = a'$ and $b = b'$. Then we can define the *product* of two sets by

$$A \times B = \{(a, b) \, : \, a \in A, b \in B\}.$$

Note that $A \times B \subset \mathcal{P}(\mathcal{P}(A \cup B))$, so it is a set.

**Relations.** A *relation* $R$ between $A$ and $B$ is a subset $R \subset A \times B$. It has a *domain* and *range*.

A relation can be visualized as a directed graph with vertices $A \cup B$ and with an edge from $a$ to $b$ exactly when $(a, b) \in \mathbb{R}$.

Examples: an equivalence relation is a subset of $A \times A$ with certain properties. The relation $i < j$ on $\mathbb{Z}$. The relation $b|a$ on $\{1, 2, \ldots, 10\}$.

**Functions.** A *function* $f : A \to B$ is a relation between $A$ and $B$ such that for each $a \in A$, there is a unique $b$ such that $(a,b) \in f$. We write this as $b = f(a)$. Functions are also called *maps*.

The set of all $f : A \to B$ is denoted $B^A$. Why? How many elements does $3^5$ have? (Answer: 243.)

A function can be *injective* and/or *surjective*. It is *bijective* if both.

Composition of maps: $f \circ g$. If $f : A \to B$ is bijective, then there is a unique map $g : B \to A$ such that $g \circ f(x) = x \,\forall x \in A$.

Examples: $f(n) = n^2$ is injective on $\mathbb{N}$, but not on $\mathbb{Z}$. It is surjective in neither case. The function $\sin : \mathbb{R} \to [-1,1]$ is surjective but not injective. Its restriction, $\sin : [-\pi/2, \pi/2] \to [0,1]$, is bijective. Its restriction, $\sin : [0,1] \to [-1,1]$, is injective but not surjective.

**Set theory as a programming language.** The point of the definitions of $\mathbb{N}$ and $(a,b)$ is not so much that they are natural or canonical, but that they work. In other words set theory provides a very simple language in which the rest of mathematics can be *implemented*.

There is a natural bijection between $A \times A$ and $A^2$.

There is a natural bijection between $\mathcal{P}(A)$ and $2^A$.

**$\mathcal{P}(X)$ as an algebra.** If we define $A \cdot B = A \cap B$ and $A + B = (A \cup B) - (A \cap B)$, then $\mathcal{P}(X)$ becomes a ring. The identity elements are $\emptyset$ and $X$. This rings is nothing but the ring of map $f : X \to \mathbb{Z}/2$.

Note that $A + A = 0$. Thus this ring is also an *algebra* over the field $\mathbb{F}_2$.

We let

$$\chi : \mathcal{P}(X) \to 2^X$$

denote the map that sends $A$ to its *indicator function* $\chi_A$ which is 1 on $A$ and otherwise zero.

**Functions, unions and intersections.** Let $f : X \to Y$ be a function. We set $f(A) = \{f(a) : a \in A\}$. In this way we obtain a map $f : \mathcal{P}(X) \to \mathcal{P}(Y)$.

Is this map a ring homomorphism?

In general, $f(A \cap B) \neq f(A) \cap f(B)$. We only have $f(A \cap B) \subset f(A) \cap f(B)$. However, if $f$ is *injective*, then equality holds. We always have, however, $f(A \cup B) = f(A) \cup f(B)$.

If $f : A \to B$ is a function, for any subset $X \subset B$ we define $f^{-1}(X) = \{a \in A : f(a) \in X\}$. Thus we have $f^{-1} : \mathcal{P}(B) \to \mathcal{P}(A)$. This map preserves intersection and unions: e.g. $f^{-1}(X \cap Y) = f^{-1}(X) \cap f^{-1}(Y)$.

Abusing notation, one also writes $f^{-1}(b)$ for $f^{-1}(\{b\})$.

**The ubiquity of $f^{-1}$.** The fact that $f^{-1}$ preserves set-theoretic operations means that a good theory of maps often turns on properties of $f^{-1}$ rather than $f$.

For example, we will see that a continuous function can be defined as one such that $f^{-1}(U)$ is open for every open set $U$. Similarly a measurable function is one such that $f^{-1}(U)$ is measurable for every open set $U$.

Another reason for this ubiquity is that functions naturally pull back. In particular, the indicator functions satisfy

$$\chi_A \circ f = f^*(\chi_A) = \chi_{f^{-1}(A)}.$$

This pullback operation preserves the algebra structure of the space of functions. In differential topology, we find that forms pullback as well, that $d(f^*\omega) = f^*(d\omega)$, etc.

**Cardinality and the Axiom of Choice.** We say sets $A$ and $B$ have the same *cardinality* if there is a bijection between $A$ and $B$. We will write this relation as $|A| = |B|$. It is an equivalence relation.

**Theorem 2.1** $|\mathbb{N}| \neq |\mathbb{R}|$.

**Proof.** Suppose $x(n)$ is a list of all real numbers, and write their fractional parts as

$$\{x_n\} = 0.x_1(n)x_2(n)\ldots$$

in base 10. Now choose any sequence of digits $y_i$ with $y_n(n) \neq x_n(n)$. We can also arrange that $y_i$ is non-repeating. Then

$$y = 0.y_1y_2y_3\ldots$$

disagrees with $x_n$ in its $n$th digit, so it is not on the list. ∎

**Theorem 2.2 (Cantor)** . *$A$ and $\mathcal{P}(A)$ do not have the same cardinality.*

**Proof.** Given $f : A \to \mathcal{P}(A)$, let $B = \{a : a \notin f(a)\}$. Suppose $B = f(a)$. Then $a \in B$ iff $a \notin B$. This is a contradiction so $f$ does not exist. ∎

If we think of $\mathcal{P}(\mathbb{N}) \cong 2^{\mathbb{N}}$ as sequences $(a_i)$ of binary digits, then the proof that $|\mathbb{N}| \neq |\mathcal{P}(\mathbb{N})|$ is almost the same as digit diagonalization.

**Corollary 2.3** *There are many different sizes of infinity.*

**Finite sets.** A set $A$ is *finite* iff there is a bijection $f : A \to n$ for some $n \in \mathbb{N}$. That is, $A$ is finite iff $|A| = |n|$ for some $n \in \mathbb{N}$.

**Theorem 2.4 (Pigeon-hole principle)** *If $A$ is finite, then any injective map $f : A \to A$ is surjective.*

**Proof.** By induction on $|A| = n$. ∎

Application: for any prime $p$ and $a \neq 0$, there is an integer $b$ such that $ab = 1 \bmod p$. (I.e. $b = 1/a$). Proof: The map $b \mapsto ab$ is $1-1$ on $\mathbb{Z}/p$, so it is onto. Example: $1/10 = 12 \bmod 17$; in fact $10 * 12 = 120 = 7 * 17 + 1$.

**Infinite sets.** A set is *infinite* iff it is not finite.

**Theorem 2.5** $\mathbb{N}$ *is infinite.*

**Proof.** Otherwise, there would for some $n$ be an injective map $\mathbb{N} \hookrightarrow n$, and hence an injective map $n+1 \hookrightarrow n$. This contradicts the pigeon–hole principle. ∎

**Axiom VII (The Axiom of Choice).** For any set $A$ there is a function $c : \mathcal{P}(A) - \{\emptyset\} \to A$, such that $c(B) \in B$ for all $B \subset A$.

In concrete cases it is possible to find explicit choice functions. For the natural numbers, we can let $c(A) = \min(A)$. For the rational numbers, we can take this 'simplest' rational number $x = p/q$ in $A$, say with $q$ minimal, $|p|$ minimal given $q$, and with $x \geq 0$ to break ties.

**The smallest infinite set.** Here is typical use of the Axiom of Choice.

**Theorem 2.6** *$A$ is infinite iff there is an injective map $f : \mathbb{N} \to A$.*

**Proof.** If $A$ is finite then any subset of $A$ is finite, so there is no injection of $\mathbb{N}$ into $A$.

Now assume $A$ is infinite; we will construct $f$. Pick some $a \in A$. Then define, by induction, $f(0) = a$ and $f(n+1) = c(A - \{f(0), \ldots, f(n)\})$. The resulting map is injective by construction. ∎

**Cantor's definition of infinity.** A set $A$ is infinite iff there exists a map $f : A \to A$ which is injective but not surjective.

The proof uses the Axiom of Choice: first embed $\mathbb{N}$ into $A$, and then use $x \mapsto x + 1$ on $\mathbb{N}$.

**Hilbert's hotel.** The set $\mathbb{N}$ (or any infinite set) serves to illustrate *Hilbert's hotel*. The hotel is full, and yet but just shuffling the residents around we can create an empty room.

Note that there is a bijection between $\mathbb{N}$ and *any* infinite subset of $\mathbb{N}$, such that the odd numbers or the squares.

There was once a University with a long line of offices had become a little top–heavy: the professors could only occupy the offices with numbers $1, 4, 9, 16, \ldots, n^2 \ldots$ because the rest were taken up by the Deans. Outraged, the president required that each professor be assigned his own Dean, and the rest fired. The Dean in office $n$ was assigned to the professor in room $n^2$, and now the Dean's were fully employed as personal assistants to professors.

No one had to be fired. In fact the professors in offices 16, 81, 256, ... were still left without assistants, so more Deans were hired.

**Other applications of AC.** Every vector space has a basis. The Hahn–Banach theorem. Every set can be well–ordered. Choice of coset representatives for $G/H$. Existence of non-measurable sets.

**The Banach-Tarski paradox.** As a consequence of AC, you can cut a grapefruit into 5 pieces and reassemble them by rigid motions to form 2 grapefruits. (Now you've gone too far.)

**Relative size.** It is natural to say $|A| \geq |B|$ if there is a surjective map from $A$ to $B$. But it is equally natural to require that there is an injective map from $B$ to $A$. The result above is used to show these two definitions are equivalent.

Small point: if $|B| = 0$ then the surjective definition does not work.

Let us say $|A| \leq |B|$ if

(1) there is an injection $f : A \hookrightarrow B$; or
(2) there is a surjection $g : B \twoheadrightarrow A$, or $A = \emptyset$.

**Theorem 2.7** *(1) and (2) are equivalent.*

**Proof.** Given the inclusion $f$ we obtain from $f^{-1}$ a surjection from $f(A)$ back to $A$, which we can extend to the rest of $B$ as a constant map so long as $A \neq \emptyset$. Conversely, using the Axiom of Choice, we take $f$ to be a section of $g$, i.e. set $f(a) = c(g^{-1}(\{a\}))$. ■

**Theorem 2.8 (Schröder-Bernstein)** *If* $|A| \leq |B|$ *and* $|B| \leq |A|$ *then* $|A| = |B|$.

**Proof.** We will assume $A$ and $B$ are disjoint — this can always be achieved, if necessary, by replacing $A, B$ with $A \times \{0\}, B \times \{1\}$.

Suppose we have injections $f : A \to B$ and $g : B \to A$. Then we obtain an injection

$$F = f \cup g : A \cup B \to A \cup B.$$

To clarify the proof, say $F(x)$ is the *child* of $x$, and $x$ is the *parent* of $F(x)$. Since $F$ is injective, a child can have only one parent, and every element of $A \cup B$ is a parent. However some parents are no-one's child; let us call them *godfathers.*

For any $x \in A \cup B$, either $x$ is descended from a unique godfather (possibly $x$ itself), or $x$ has no godfather; it has an infinite line of ancestors (or $x$ is descended from itself.)

Now partition $A$ into 3 pieces, $A_0$, $A_A$ and $A_B$. $A_0$ is the elements $x \in A$ with no godfather; $A_A$ consists of those $x$ whose godfather is in $A$; and $A_B$ is those whose godfather is in $B$. Similarly define $B_0, B_A, B_B$.

There is a bijection $A_0 \leftrightarrow B_0$ defined by sending $a$ to its child $F(a)$. It is injective because $F$ is, and it is surjective because every $x \in B_0$ has a parent, which must lie in $A_0$.

There is a bijection $A_A \leftrightarrow B_A$ defined by sending each $a \in A_A$ to its child $F(a)$. The inverse map sends children to their parents. There are no godfathers in $B_A$, so the inverse is well-defined.

Similarly there is a bijection $A_B \leftrightarrow B_B$, sending $a \in A$ to its parent in $B_B$. Putting these three bijections together shows $|A| = |B|$. ■

**Countable sets.** We say $A$ is countable if $|A| \leq |\mathbb{N}|$. Finite sets are countable.

**Theorem 2.9** *If $A$ is countable and infinite, then* $|A| = |\mathbb{N}|$.

**Proof.** Infinite implies $|\mathbb{N}| \leq |A|$, and countable implies $|A| \leq |\mathbb{N}|$; apply SB. ■

**Theorem 2.10** $\mathbb{N}^2$ *is countable.*

**Proof.** Define a bijection $f : \mathbb{N} \to \mathbb{N}^2$ by $f(n) = (a, b)$ where $n + 1 = 2^a(2b + 1)$. ■

**Corollary 2.11** *A countable union of countable sets is countable.*

**Proof.** If $X = \bigcup A_i$ we can send the $j$th element of $A_i$ to $(i, j) \in \mathbb{N} \times \mathbb{N}$. ■

**Examples.**

1. The set of things that can be described in words is countable. Thus most real numbers have no names.
2. The integers $\mathbb{Z}$ can be constructed from $2 \times \mathbb{N}$; they satisfy $|\mathbb{Z}| = |\mathbb{N}|$.
3. The rationals $\mathbb{Q}$ are $\mathbb{Q} = \mathbb{Z} \times \mathbb{Z}^*/\sim$, where $(a, b) \sim (c, d)$ if $ad - bc = 0$. The $\mathbb{Q}$ is countable.
4. The ring of polynomials $\mathbb{Z}[x]$ is countable.

**Corollary 2.12** *Most real numbers are transcendental.*

**Warning.** While a countable sum of countable sets is countable, the same is not true for products. Indeed, $2^{\mathbb{N}}$ is a countable product of *finite* sets, but it naturally isomorphic to $\mathcal{P}(\mathbb{N})$.

**Theorem 2.13** $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$.

**Proof.** We can associate to each subset $A \subset \mathbb{N}$ a unique real number defined in base 2 by

$$x_A = x_0.x_1x_2x_3\ldots = \sum_{n \in A} 2^{-n},$$

where $x_n = 1$ if $n \in A$ and zero otherwise. Conversely, a real number $x$ is uniquely determined by the set

$$A_x = \{y \in \mathbb{Q} \,:\, y < x\} \subset \mathbb{Q}.$$

Thus $|\mathcal{P}(\mathbb{N})| \le |\mathbb{R}| \le |\mathcal{P}(\mathbb{Q})| = |\mathcal{P}(\mathbb{N})|$. ■

**Theorem 2.14** $|\mathbb{R}^{\mathbb{R}}| = |\mathcal{P}(\mathcal{P}(\mathbb{N}))|$. *Thus the functions on* $\mathbb{R}$ *represent the third kind of infinity.*

**Proof.** First notice that we have the easy inequality:

$$|\mathcal{P}(\mathcal{P}(\mathbb{N}))| = |\mathcal{P}(\mathbb{R})| = |2^{\mathbb{R}}| \leq |\mathbb{R}^{\mathbb{R}}|.$$

On the other hand, we can construct an injection $i : \mathbb{R}^2 \to \mathbb{R}$ by interleaving decimal digits. Since a function $f : \mathbb{R} \to \mathbb{R}$ is a special kind of relation $f \subset \mathbb{R} \times \mathbb{R}$, we then have:

$$|\mathbb{R}^{\mathbb{R}}| \leq |\mathcal{P}(\mathbb{R}^2)| = |\mathcal{P}(\mathbb{R})|,$$

and so $|\mathbb{R}^{\mathbb{R}}| = |\mathcal{P}(\mathcal{P}(\mathbb{N}))|$ by the Schröder-Bernstein theorem. ∎

Another way to think of the fact that $|\mathbb{R}^2| = |\mathbb{R}|$ is that $|\mathbb{N} \times 2| = |\mathbb{N}|$, and thus

$$|\mathbb{R}^2| = |\mathcal{P}(\mathbb{N}) \times \mathcal{P}(\mathbb{N})| = |\mathcal{P}(\mathbb{N} \times 2)| = |\mathcal{P}(\mathbb{N})| = |\mathbb{R}|.$$

**The continuum hypothesis.** Using the Axiom of Choice, one can prove that for any two sets $A$ and $B$, $|A| \leq |B|$ or $|B| \leq |A|$.

Is there a set $A$ such that $|\mathbb{N}| < |A| < |\mathbb{R}|$? It is now known that this question *cannot be answered* using the axioms of set theory (assuming these axioms are themselves consistent). Some logicians have argued that CH is obviously false (Cohen, Woodin), while others have argued that it must be true (Woodin).

# 3 Topology

This section describes the basic definitions and constructions in topology, together with lots of examples.

**Topology.** A *topological space* is a set $X$ equipped with a distinguished collection of *open sets* $\mathcal{T} \subset X$. That is, $U \subset X$ is *open* iff $U \in \mathcal{T}$. We require that:

1. $\emptyset, X \in \mathcal{T}$;
2. If $U, V \in \mathcal{T}$ then $U \cap V \in \mathcal{T}$; and

3. If $\mathcal{A} \subset \mathcal{T}$ then $\bigcup \mathcal{A} \subset \mathcal{T}$.

More informally, any union of open sets is again open, and the intersection of any *finite* collection of open sets is again open.

Usually we will refer to $(X, \mathcal{T})$ simply as a *space* (as opposed to a set), and denote it by $X$.

**Closed sets.** A set $F$ is closed (*fermé*) if its complement $U = X - F$ is open. The axioms above show the collection of closed sets in preserved by arbitrary intersections and finite unions. One can also specify a topology by giving the collection of all closed sets.

**Euclidean space and metric spaces: Paris and Manhattan.** The natural topology $\mathcal{T}$ on a metric space $(X, d)$ is defined by saying $U$ is open iff for all $x \in U$ there is an $r > 0$ such that the ball $B = B(x, r)$ satisfies $x \in B \subset U$.

In other words, there should be some space around each point, like on the wide *open* range.

If you are not so familiar with general metric spaces, you can always think about $\mathbb{R}^n$ with $d(x, y) = |x - y|$. To distinguish metrics and topology, consider the Manhattan metric on $\mathbb{R}^n$: $d(0, (x, y)) = |x| + |y|$. This gives the same topological space.

Exercise. The Paris metric on $\mathbb{R}^2$ is defined by $d(x, y)$ is the minimum length of a path from $x$ to $y$ passing through 0, *unless* $x$ and $y$ happen to lie on the same ray.

Exercise: How is the *topology* of this space essentially different from the usual topology on $\mathbb{R}^2$?

**Bases.** This example suggests more generally defining $\mathcal{T}$ using a *basis* $\mathcal{B} \subset \mathcal{P}(X)$. A basis is a collection of sets such that

1. $X = \bigcup \mathcal{B}$; and
2. If $B_1$ and $B_2$ in $\mathcal{B}$ overlap at $x$, then there is a $B_3 \in \mathcal{B}$ such that $x \in B_3 \subset B_1 \cap B_2$.

We then say $U \subset X$ is *open* if it is a union of elements of $\mathcal{B}$. It is easy to verify that the set $\mathcal{T}$ of all such $U$ forms a topology. Closure under unions follows from the definitions, and closure under finite intersections follows from (2).

The topology $\mathcal{T}$ is simply the smallest topology which contains $\mathcal{B}$.

Clearly the collection $\mathcal{B}$ of all open balls $B(x, r)$ form a basis for the usual topology on a metric space.

**How complex can an open set be?** The interval $(a, b)$ form a basis for the topology of $\mathbb{R}$. In fact, every open set can be written as a union of *disjoint intervals* (exercise).

We may need infinitely many intervals to describe $U$: consider, for example, $U = \bigcup_{\mathbb{N}}(n, n+1)$, $\bigcup_{n>0}(1/(n+1), 1/n)$. Here at least the intervals appear in order.

That is not always the case! There exist open sets such that between any two disjoint intervals $I, J \subset U$ there is another interval $K$. A good example is provided by

$$U = \{x \in (0, 1) \ : \ x \text{ has a robust 5 in its decimal expansion}\}.$$

To be more precise, we define $U_0 = (0.5, 0.6)$ as the set of numbers that have a robust 5 as their first digit after the decimal point. We have excluded 0.5 and $0.599999\ldots$, since these are not robust — a small perturbation gets rid of the 5.

Then we define $U_n = \{x \in (0, 1) \ : \ \{10^n x\} \in U_0\}$. Here $\{x\} = x - [x]$ is the fractional part of $x$. These are the numbers with a robust 5 in the $n$th decimal place.

Finally we define $U = \bigcup U_n$.

Then $U_1$ contains the new interval $(0.05, 0.06)$, $(0.15, 0.16)$, $(0.25, 0.26)$, etc. as well as the intervals like $(0.55, 0.56)$ that are already contained in $U_0$. Each $U_{n+1}$ contains many new intervals like this. The complement of $U$ in $[0, 1]$ is a *Cantor set*.

**Continuous functions.** A function $f : X \to Y$ is *continuous* if $f^{-1}(U)$ is open for all open sets $U \subset Y$.

Intuitively, continuity means:

*$f(x)$ is close to $f(a)$ whenever $x$ is close enough to $a$.*

To say '$f(x)$ is close to $f(a)$' is to say that $f(x)$ lies in a neighborhood $V$ of $f(a)$. Let $U = f^{-1}(V)$. Then $f(x)$ is 'close to $f(a)$' iff $x \in U$. *When $f$ is continuous*, $U$ is a neighborhood of $a$, and when $x$ is in the neighborhood $U$ of $a$, it is 'close enough to $a$' that $f(x) \in V$.

**Theorem 3.1** *The composition of two continuous maps is continuous.*

**Theorem 3.2** *A map $f : X \to Y$ is continuous iff $f^{-1}(B)$ is open for all element $B$ in a basis for the topology on $Y$.*

**Proof.** For any open set $U = \bigcup B_i$, the preimage $f^{-1}(U) = \bigcup f^{-1}(B_i)$ is a union of open sets, hence open. ■

**Example.** To verify a map $f : X \to \mathbb{R}$ is continuous, it suffices to verify that $f^{-1}(a, b)$ is open for all $(a, b)$.

**Homeomorphism.** A bijection between spaces, $f : X \to Y$, is a *homeomorphism* if both $f$ and $f^{-1}$ are continuous. This is the same as saying that $U$ is open $\iff$ $f(U)$ is open.

Homeomorphic spaces are *topologically the same.*

Warning: it is tempting to think that a bijective continuous map is always a homeomorphism. This is not true! For example, let $X = \{0\} \cup \{01, 1/2, 1/3, \ldots\}$. Then the map $f : \mathbb{N} \to X$ given by $f(0) = 0$, $f(n) = 1/n$ for $n > 0$, is continuous and bijective, but $X$ is not homeomorphic to $\mathbb{N}$.

We will later see that a continuous bijective map between *compact Hausdorff spaces* is always a homeomorphism.

**The reals and $(0, 1)$.** A metric space is *bounded* if $\sup d(x, y) < \infty$. One of the most basic differences between metric spaces and topological spaces is that the overall size of a space is hard to detect topologically. As an important example we note:

The real numbers $\mathbb{R}$ are homeomorphic to the open interval $(0, 1)$.

A fancy way to see this is to first note that all bounded intervals $(a, b)$ are related by linear maps, and then note that

$$\tan : (-\pi/2, \pi/2) \to \mathbb{R}$$

is a homeomorphism. Many other functions will do the trick.

**Real-valued functions.** We let $C(X, Y)$ denote the set of all continuous map $f : X \to Y$. We let $C(X) = C(X, \mathbb{R})$ denote the algebra of real–valued continuous functions on $X$.

It is easy to show that if $f$ and $g$ are continuous, then so are the sum and product $f + g$, $fg$. Thus $C(X)$ is an *algebra.* It always contains the constant functions $\mathbb{R}$.

In a metric space there are always lots of interesting functions in $C(X)$, since $f(x) = d(x, x_0)$ is continuous. In a general topological space, there might be lots or there might be none. One of our tasks is to *construct* lots of elements of $C(X)$, under suitable hypotheses on $X$.

**Functions and topology.** If we broaden our test targets beyond $\mathbb{R}$, the space of continuous functions on $X$ uniquely determines its topology.

As a simple example, let $Z = \{0, 1\}$ with the topology where $\{1\}$ is open but not $\{0\}$ is not. Then $A$ is open iff $\chi_A$ is continuous. This shows:

**Theorem 3.3** *The topology on $X$ is uniquely determined by the set of continuous functions $C(X, Z)$.*

**Limits.** A *neighborhood* of $b \in X$ is just an open set $U$ with $b \in U$. We say $a_n \to b$ if for any neighborhood $U$ of $b$, we have $a_n \in U$ for all $n \gg 0$.

**Two basic topologies.** In the *discrete topology*, all sets are open, and all functions are continuous, so $C(X) = \mathbb{R}^X$. In the *trivial topology*, only $X$ and $\emptyset$ are open, so $C(X) = \mathbb{R}$.

**Cofinite topology.** A slightly more interesting topology is the *cofinite* topology. In this topology, $A \subset X$ is closed iff $|A| < \infty$ or $A = X$.

If $X$ is finite, then the cofinite topology is just the discrete topology, but otherwise it is different.

In the cofinite topology, if $a_n$ is an infinite sequence, then $a_n \to b$ for all $b \in X$. In particular limits are not unique!

**Example: The lower–limit topology: $\mathbb{R}_\ell$ .** An interesting topological space is given by $\mathbb{R}_\ell$, the real numbers with the intervals $[a, b)$ as a basis of open sets. In this topology, $a_n \to a$ iff $a_n$ converges to $a$ in the traditional sense *and* $a_n \geq a$ for all $n \gg 0$.

A function $f : \mathbb{R}_\ell \to \mathbb{R}$ is continuous iff $f(a_n) \to f(b)$ whenever $a_n \to b$ *from above.*

**Uphill topology.** We can think of $\mathbb{R}_\ell$ as the *uphill* topology. It requires huge effort to climb the hill, so points cannot be approached from below. However it is easy to slide down to them. This lack of symmetry is not consistent with a metric, since the latter satisfies $d(x, y) = d(y, x)$. In fact, $\mathbb{R}_\ell$ is not metrizable.

**Comparison of $\mathbb{R}_\ell$ with $\mathbb{R}$.** Since $[a, b)$ is *not* generally open in $\mathbb{R}$, the identity map $i : \mathbb{R} \to \mathbb{R}_\ell$ is not continuous. (I.e. $\mathbb{R}_\ell$ is not swept out by a path.) On the other hand, $i : \mathbb{R}_\ell \to \mathbb{R}$ *is continuous.*

**Order topology.** If $X$ is given a total ordering, then it becomes a topological space $X_o$ by taking the intervals $(a, b) = \{x : a < x < b\}$ as a basis.

**The product topology.** Let $X$ and $Y$ be topological spaces. The *product topology* on $X \times Y$ is defined by taking as a basis the sets of the form $U \times V$, where $U$ and $V$ are open sets in $X$ and $Y$.

**Theorem 3.4** *(1) The product topology is the smallest topology such that the projections of $X \times Y$ to $X$ and $Y$ are continuous.*

*(2) A function $f : Z \to X \times Y$, given by $f(z) = (f_1(z), f_2(z))$, is continuous iff each of its coordinates $f_1$ and $f_2$ are continuous.*

**Proof.** (2) It suffices to check continuity using the basic open sets $U \times V$ in $X \times Y$. For these, $f^{-1}(U \times V) = f_1^{-1}(U) \cap f_2^{-1}(V)$ is an intersection of two open sets, and hence open. ∎

**Topologies on the square.** The product topology on $I \times I$ coincides with the usual metric topology. The lexicographical or dictionary order topology gives a weirder space, $I_o^2$, the *ordered square*.

To understand this topology, we must see what a small interval around a point $(a, b)$ looks like. If $0 < b < 1$, then $\{a\} \times (b', b'')$ is a typical small interval around $(a, b)$. So open vertical segments qualify as open set, and any union of these is also open. This shows any 'usual' open subset of the interior of $I^2$ is also open in $I_o^2$.

But what is a neighborhood of $(a, 1)$ in $I_o^2$? If $0 < a < 1$ then it must contain some whole vertical segment $a' \times I$! In particular, a small neighborhood of $(a, 1)$ in the usual topology is *not* open in the order topology.

Think of this as convergence in a dictionary. Most of the time, if a sequence of words satisfy $w_n \to w$, then eventually the first letter of $w_n$ agree with that of $w$.

An exception happens, for example, when $w =$ syzyzgy is the last word in the $S$ section. Then any neighborhood of $w$ contains some word beginning with $t$. See also *azygous, Byzantine, Czech, dysuria, Ezra, fylfot*, etc.

**The subspace topology.** If $A \subset X$ and $X$ is a topological space, the *subspace topology* is defined by

$$\mathcal{T}_A = \{U \cap A : U \in \mathcal{T}_X\}.$$

This is the unique topology so that if $f : Z \to A$ is continuous $\iff$ $\iota \circ f : A \to X$ is continuous, where $\iota : A \to X$ is the inclusion.

**Interval example.** Let $I = [0, 1] \subset \mathbb{R}$ with the subspace topology. Then $[0, a)$ and $(b, 1]$ are open sets in $I$, even though they are not open in $\mathbb{R}$.

**Order example.** The ordered square $I_o^2$ is a subset of the ordered plane $\mathbb{R}_o^2$. Is the order topology on the square the same as the subspace topology?

Weirdly, the answer is no (see Munkres, Fig. 16.1).

**Axes example.** Let $X = \mathbb{R} \cup i\mathbb{R} \subset \mathbb{C}$, with the subspace topology. When is an interval $(a, b)$ along the real axis open in $X$? Only when $0 \notin (a, b)$!

**Topology as a language.** We have already seen how to define open and closed sets.

*Closure.* The closure $\overline{A}$ of a set $A \subset X$ is simply the smallest closed set containing $A$. This exists because the intersection of closed sets is closed.

*Interior.* Similarly, the interior of $A$, denoted $\operatorname{int}(A)$, is the largest open set contained in $A$.

*Boundary.* The boundary of $A$ is their difference: $\partial A = \overline{A} - \operatorname{int} A$. A point $x$ lies in the boundary of $A$ iff ever neighborhood of $x$ meets both $A$ and $X - A$.

*Density.* We say $A \subset X$ is dense if $\overline{A} = X$.

Example: the boundary of the open unit ball $B$ in $\mathbb{R}^2$ is $S^1$. The same is true of the closed unit ball. The boundary of the set of points $A$ in the unit ball $B$ which happen to have rational coordinates is the whole ball $B$. The interior of $A$ is empty, and $A$ is dense in $B$ (in the subspace topology).

**Limit points.** An essential notion in calculus is the idea of limit. But it is usually expressed in terms of sequences. How can it be expressed just using open sets?

Answer: we say $x$ is a *limit point* of $A$ if every neighborhood of $x$ meets $A - \{x\}$. In a metric space, this would imply that there is a sequence $a_n \in A$ with $a_n \neq x$ for all $n$ and $a_n \to x$.

**Theorem 3.5** *The closure of $A$ is the union of $A$ and its limit points.*

**Proof.** Suppose $x \notin A$. If $x$ is not a limit point, then there is a neighborhood $U$ of $X$ which does not meet $A$. Then $\overline{A} \cap U = \emptyset$ as well, so $x \notin \overline{A}$. Conversely, if $x \notin \overline{A}$ then $U = X - \overline{A}$ is an open neighborhood of $x$ disjoint from $A$, so $x$ is not a limit point. ■

A point $x \in A$ is *isolated* if there is a neighborhood $U$ of $x$ such that $U \cap A = \{x\}$. Exercise: every point of a closed set $A$ is either a limit point or an isolated point. The set of isolated points in $A$ form an open subset of $A$, and the set of limit points form a closed subset of $A$.

**Examples: derived sets.** The set of limit points $A'$ of a closed set $A \subset \mathbb{R}$ is called the *derived* set of $A$. It is obtained by just throwing out the isolated points. For example, if $A = \{0, 1, 1/2, 1/3, \ldots\}$ then $A' = \{0\}$ and $A'' = \emptyset$. On the other hand, if $A = [0, 1]$ then $A' = A$.

Can $A' = A$ if $A$ does not contain an interval? The answer is yes. The *Cantor set* $K \subset [0, 1]$ has $K' = K$ even though its interior is empty.

**Derived sets and the invention of ordinals*.** Another remarkable fact, which we will not prove, is that a countable closed set $A \subset [0, 1]$ always has an isolated point. Thus $A \neq A'$. We can then define $A_0 = A$ and $A_{n+1} = A'_n$. It may happen that all these sets are nonempty, and then $A_\infty = \bigcap A_n \neq \emptyset$ (by compactness). We can then continue by setting $A_{\infty+1} = A'_\infty$, and so in. This led Cantor to the invention of ordinals and much of the set theory we have discussed.

Obviously $A_1 = \{0, 1, 1/2, 1/3, \ldots\}$ is a countable closed set with $A'_1$ nonempty. If we stick together a sequence of copies of $A_1$ accumulating on 0, we obtain a set $A_2$ with $A''_2$ nonempty. Similarly we can inductively construct closed, countable sets $A_n$ so that $A_n^{(n)} \neq \emptyset$. Then we can strict together $(A_1, A_2, \ldots)$ to create a set $A_\omega$ such that

$$A_\omega^{(\omega)} = \bigcap_{i=1}^{\infty} A_\omega^{(i)} \neq \emptyset.$$

Continuing in this way, one can construct for any countable ordinal a countable closed set $A_\alpha$ such that $A_\alpha^{(\alpha)} \neq \emptyset$.

Cantor was motivated by a problem in Fourier series. A closed set $A \subset [0, 2\pi]$ is a *set of uniqueness* if the only way a Fourier series $\sum a_n e^{int}$ can converge to zero at each point of $[0, 2\pi] - A$ is if all its coefficients $a_n$ are zero. Using derivations, Cantor showed any countable closed set is a set of uniqueness. On the other hand, the linear middle $\xi$ Cantor set $K(\xi)$ is sometimes a set of uniqueness, sometimes not — e.g. the case where $1/\xi$ is a Pisot number is problematic. See Salem, *Algebraic numbers and Fourier analysis.*

**Example: sequences are not enough!** Let $X = \mathbb{R}$ with the topology $\mathcal{T}$ where the closed sets are just $\mathbb{R}$ and the *countable* sets $A \subset \mathbb{R}$. Take any uncountable set in $\mathbb{R}$, say $B = (0,1)$. Then $\overline{B} = \mathbb{R}$. But there is no sequence $a_n \in B$ with $a_n \to 2$! In fact, any sequence in $B$ is already closed!

**Hausdorff spaces.** Suppose $a_n \to x \in X$. Can it also be true that $a_n \to y$ in $X$, with $x \neq y$?

The answer is yes: in fact, if $X$ has the trivial topology, then every sequence converges to every point, since the only nonempty open set is $X$.

We say $X$ is *Hausdorff* if any pair of distinct points have disjoint neighborhoods. In *this* case limits are unique!

The Hausdorff condition is one of the *Trennungsaxioms*; it is traditionally denoted $T_2$. The axiom $T_1$ says points are closed. It is easy to see that $T_2 \implies T_1$.

Example. If $X$ is infinite, the cofinite topology on $X$ is not Hausdorff. But it *is* $T_1$.

This type of example comes up in 'real life': if $X$ is a variety over $\mathbb{C}$ of dimension $> 0$, then the Zariski topology on $X$ is not Hausdorff.

**Theorem 3.6** *The product of two Hausdorff spaces is again Hausdorff. A subspace of a Hausdorff space is Hausdorff.*

**Continuity in products.** It is tempting to think the a function $f(x,y)$ on $X \times Y$ is continuous if $y \mapsto f(x_0, y)$ and $x \mapsto f(x, y_0)$ are continuous for each fixed $x_0, y_0$. But this is not the case.

The traditional counterexample on $\mathbb{R}^2$,

$$F(x,y) = xy/(x^2 + y^2)$$

with $F(0,0) = 0$, is verifiable but mysterious. Here is a better example. Consider instead a map from $\mathbb{R}^2$ into $S^1 = \mathbb{R}/2\pi\mathbb{Z}$. As a first attempt we take $f(x,y) = \theta$ in polar coordinates. But this is not continuous on lines through the origin.

To fix that, we observe that $2\theta$ *is* continuous, since $\theta$ jumps by $\pi$ when it crosses the origin. Now if we set $f(0,0) = 0$, then $f(x,y)$ is continuous on horizontal lines. And if we set $f(0,0) = \pi$, it is continuous on vertical lines. To complete the example, we just set

$$f(x,y) = 4\theta(x,y) \quad \text{and} \quad f(0,0) = 0.$$

Then $f$ is continuous on horizontal and vertical lines, but its limit is different along lines through $(0,0)$ with other slopes, so it is not continuous on $\mathbb{R}^2$.

For a real-valued function, one can take $\cos f(x,y)$ or $\sin f(x,y)$. The same construction would work with *any* function $f(x,y) = F(\theta)$, provided $F : S^1 \to \mathbb{R}$ is continuous, $F(\theta) = 0$ for $\theta = 0, \pi/2, \pi, 3\pi/2$, and $F(\theta)$ is not identically zero.

The 'traditional' counterexample is proportional to

$$f(x,y) = \sin(2\theta) = \frac{\operatorname{Im} z^2}{|z|^2} = \frac{2xy}{x^2+y^2}.$$

More generally, if we take any two homogeneous polynomials in $x, y$ of the same degree, $P$ and $Q$, with $Q(x,y) \neq 0$ for $(x,y) \neq 0$, then we can set

$$f(x,y) = \frac{P(x,y)}{Q(x,y)}$$

away from the origin. Then, if $P(1,0) = P(0,1) = 0$, we can extend $f$ by $f(0,0) = 0$ to get a function continuous in $x$ and $y$ individually. Finally, if $P(x,y)$ is not identically zero, $f(x,y)$ will not be continuous.

**Normality.** Once points are closed, we can impose the stronger *normality* condition ($T_4$) that any two disjoint *closed* sets, $A, B \subset X$, can be engulfed by disjoint *open sets*, $U \supset A$, $V \supset B$. We will later see that this powerful condition insures there is a continuous function $f : X \to \mathbb{R}$ such that $f|A = 1$ and $f|B = 0$. In particular, $C(X)$ is nontrivial (so long as $|X| > 1$).

**Infinite products.** Let $X_i$ be a collection of topological spaces indexed by $i \in I$. Their product $X = \prod X_i$ consists of maps $x : I \to \bigcup X_i$ such that $x_i \in X_i$.

The *product topology* on $X$ is defined by the basis of open sets

$$U = \prod U_i$$

where each $U_i \subset X_i$ is open, *and* $U_i = X_i$ for all but finitely many $i$. (Check that these basis elements satisfy $B_1 \cap B_2 \supset B_3$.)

**Theorem 3.7** *A map $f : Z \to \prod X_i$ is continuous iff each of its coordinates $f_i : Z \to X_i$ is continuous.*

**Example.** Let $2 = \{0,1\}$ be given the discrete topology. Then $2^{\mathbb{N}} \cong \mathcal{P}(\mathbb{N})$ becomes a topological space. Two subsets of $\mathbb{N}$ are close if they have the same intersection with $[0,N]$ for large $N$.

The traditional Cantor set $K \subset [0,1]$ is homeomorphic to $2^{\mathbb{N}}$, and thus provides a way of visualizing all the subsets of $\mathbb{N}$.

It turns out that for any sequence of nonempty finite sets $X_i$, each with the discrete topology, the space $\prod X_i$ is homeomorphic to the Cantor set. This is already tricky to see for $3^{\mathbb{N}}$.

**Pointwise convergence.** Fix any set $S$, and let

$$X = \mathbb{R}^S = \{\text{all functions } f : S \to \mathbb{R}\}$$

with the product topology. Then a sequence $f_n \in X$ converges to $g$ iff $f_n(s) \to g(s)$ for all $s \in S$. Thus the product topology is sometimes called, in the case $X = A^B$, the topology of *pointwise convergence*.

**The Hilbert cube.** Another nice example of an infinite product is $X = [0,1]^{\mathbb{N}}$. This is simply the space of all sequences $(x_i)$ with $0 \le x_i \le 1$. The topology is induced by a nice metric, e.g.

$$d(x,y) = \max |x_n - y_n|/2^n.$$

**Box topology.** The *box topology* on $X = \prod X_i$ is defined by using all products of open sets, $\prod U_i$, to form a base for the topology.

This topology is very strong (i.e. it has lots of open sets), so there are very few continuous maps into the product.

**Exercise.** The diagonal map $D$ of $[0,1]$ into $[0,1]^{\mathbb{N}}$ is continuous in the product topology, but discontinuous in the box topology.

(Proof: Each factor $D_i : [0,1] \to [0,1]$ is simply the identity map, which is continuous, so the product topology is fine. But in the weak topology, for any $x \in [0,1]$ we can find a sequence of open sets $U_i$ containing $x$ and nesting down to it. Then $D^{-1}(\prod U_i) = \bigcap U_i = \{x\}$ is not open, so $D$ is not continuous.)

**Topological groups.** A topological group is a set $G$ equipped with continuous product and inverse maps, $G \times G \to G$ and $G \to G$, that satisfy the usual group axioms.

The most basic topological groups are $(\mathbb{R}, +)$ and $(\mathbb{R}^*, \cdot)$. The fact that these are groups shows that $C(X)$ is an *algebra*. The fact that inversion is continuous shows that $1/f \in C(X)$ whenever $f$ has no zeros.

Other examples include the matrix group $\mathrm{GL}_n(\mathbb{R})$, which can be regarded via matrices as an open subset of $\mathbb{R}^{n^2}$. Cramer's rule shows inverse is continuous.

Here is a typical theorem about topological groups.

**Theorem 3.8** *If the origin is closed in $G$, then $G$ is Hausdorff.*

**Proof.** Since right multiplication (say) is a homeomorphism, it suffices to show that identity element $e$ and any other element $x \neq e \in G$ have disjoint neighborhoods. Since $e$ is closed, so is $x$, so there is a neighborhood $U$ of $e$ disjoint from $x$.

Now comes the key point: since multiplication is continuous, there is a smaller neighborhood $V$ of $U$ such that $V \cdot V \subset U$. We may also assume $V = V^{-1}$. It is then readily verified that $V$ and $xV$ are disjoint. ∎

# 4 Connected spaces

A space $X$ is *connected* if it cannot be written as the disjoint union of two nonempty open sets, $X = U \cup V$.

When $X$ breaks up into two open pieces $U$ and $V$, each piece is a *clopen* — it is both open and closed.

A subset $A \subset X$ is connected if it is connected in the subspace topology.

A space is *totally disconnected* if its only connected subsets are single points.

**Two basic facts about connectedness.** One is concrete and one is abstract. The concrete fact is:

**Theorem 4.1** *The interval $[0,1]$ is connected.*

**Proof.** Suppose $[0,1]$ is the union of two disjoint open sets, $U$ and $V$. We may assume $0 \in U$. Let $a = \sup\{x \in [0,1] \ : \ [0,x) \subset U\}$. Since $U$ is open, $a > 0$. If $a = 1$ we are done. Otherwise $a \notin U$, and therefore $a \in V$. But then a neighborhood of $a$ is in $V$, so we cannot have $[0,a) \subset U$. ∎

The abstract fact is:

**Theorem 4.2** *If $f : X \to Y$ is continuous and $X$ is connected, then $f(X)$ is connected.*

**Proof.** Suppose to the contrary we can write $f(X) = U \cup V$ as the union of two nonempty, disjoint open sets. Then $X = f^{-1}(U) \cup f^{-1}(V)$ does the same for $X$. ∎

**Exercise.** What does this have to do with the intermediate value theorem?

**Idempotents.** It is a beautiful fact that connectedness is detected by *idempotents* in the algebra $C(K)$, i.e. elements such that $f^2 = f$. On a connected space, the only idempotents are $f = 0$ and $f = 1$. But if $X = U \sqcup V$, then $\chi_U$ is continuous and $\chi_U^2 = \chi_U$.

**Locally constant functions.** By the same token, if $X$ is connected and $f : X \to \mathbb{R}$ is continuous and locally constant, then $f$ is constant.

**Path-connected spaces.** To show $X$ is connected, it is often easier to draw paths between points than to analyze open partitions of $X$. We say $X$ is *path connected* if for all $x, y \in X$ there is a continuous map $f : [a, b] \to X$ such that $f(a) = x$ and $f(b) = y$.

**Theorem 4.3** *A path connected space is connected.*

**Proof.** Suppose $X = U \sqcup V$ and $x \in U$. Pick a path $f : [a, b] \to X$ connecting $x$ to any other point $y$. Then $f([a, b])$ is connected, so $y \in U$. Thus $V = \emptyset$. ■

The general principle at play here is: if any two points $x, y \in X$ lie in a connected subset $A \subset X$, then $X$ is connected. The role of $A$ is played by $f([a, b])$ in the proof above.

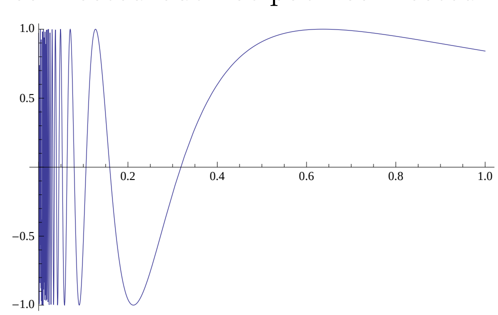
**Convex sets.** A subset $K \subset \mathbb{R}^n$ is *convex* if whenever $x, y \in K$, the straight line from $x$ to $y$ also lies in $K$. Since this straight line is a path, we find:

*Any convex set $K$ is path connected.*

Note that we do not require $K$ to be open or closed.

**Exercise.** A set $A \subset \mathbb{R}$ is connected $\iff$ it is convex $\iff$ it is path connected.

**The topologist's sine curve.** The union of the graph of $y = \sin(1/x)$ over $(0, 1]$ and the interval $[-1, 1]$ along the $y$-axis is a classic example of a closed subset of $\mathbb{R}^2$ that is connected but not path connected.



A somewhat similar set $X \subset \mathbb{R}^2$ is obtained by taking the union of the graphs of $y = 0$ and $y = 1/x$ over $[0, \infty)$. But this $X$ is disconnected. What is the difference between these two examples?

**Path connectedness in $\mathbb{R}^n$.** On the other hand, for well-behaved spaces like manifolds, connectedness and path–connectedness are the same. The proof is a typical use of the property of connectedness: we show the set of good points is both open and closed, so it must be the whole space.

**Theorem 4.4** *An open set $U \subset \mathbb{R}^n$ is connected $\iff$ it is path connected.*

**Proof.** Suppose $U$ is connected and $x \in U$. Let $V \subset U$ be the set of points that can be reached from $x$ by a polygonal path. Clearly $V$ is open. We claim it is also closed (in $U$). Indeed, if $y \in \overline{V}$ then there is a ball $B$ with $y \in B \subset U$, and a $z \in B$ that can be reach from $x$ by a polygonal path. But $B$ is convex, so $y$ can also be reached.

The converse is a general fact. ■

This proof is much easier than, say, drawing a line between $x$ and $y$ in $U$, and then explicitly modifying it so it stays in $U$.

**Example: Countable metric spaces.** A space is *totally disconnected* if the only connected sets it contains are single points.

**Theorem 4.5** *Every countable metric space $X$ is totally disconnected.*

**Proof.** Given $x \in X$, the set $D = \{d(x, y) \ : \ y \in X\}$ is countable; thus there exist $r_n \to 0$ with $r_n \not\in D$. Then $B(x, r_n)$ is both open and closed, since the sphere of radius $r_n$ about $x$ is empty. Thus the largest connected set containing $x$ is $x$ itself. ■

**The continuum.** The real numbers $\mathbb{R}$ are sometimes called the *continuum*, because they form a connected set, corresponding to the geometric notion of a line.

> *Points*
> *Have no parts or joints*
> *How then can they combine*
> *To form a line?* —J. A. Lindon.

In fact if we only consider *points with names*, these form a thin (countable subset) sprinkled along the continuum; the uncountably many remaining (fictional) points serve to connect them.

**A countable connected Hausdorff space.** (Due to Urysohn?) It is thus very curious that there *are* countable connected spaces which are Hausdorff! Of course they are not metrizable or path connected.

Let $H$ be the set of points in the closed upper half-plane with rational coordinates. Any point in $H$ is the vertex of a (possibly degenerate) equilateral triangle $T(p)$ resting on the $x$-axis. Let a neighborhood of $p$ consist of $p$ itself union a pair of intervals on the $x$-axis about the feet of $T(p)$. Since a line with slope $60^\circ$ passes through at most one rational point, the feet of $T(p)$ and $T(q)$ are disjoint if $p \neq q$ and thus $H$ is Hausdorff. On the other hand, any open set contains an interval on the $x$-axis, and the closure of such an interval consists of two bands forming an infinite copy of the letter $V$. Any two of these $V$'s intersect, so $H$ is connected. ∎

**Fans and products.** Here are two more basic properties of connectedness.

**Theorem 4.6** *A fan of connected sets is connected. That is, if $A_i \subset X$ are connected subspaces which all have a point in common, then $Y = \bigcup A_i$ is connected.*

**Proof.** Let $Y = U \sqcup V$ and suppose the common point $p$ lies in $U$. Note that $U \cap A_i$ and $V \cap A_i$ are disjoint open sets in $A_i$. Since $A_i$ is connected, and $p \in U$, we have $A_i \subset U$. Thus $Y = \bigcup A_i \subset U$ and therefore $Y$ is connected. ∎

**Corollary 4.7** *A product of connected sets $A \times B$ is connected.*

**Proof.** Suppose $A \times B = U \sqcup V$, $(a, b) \in U$, and we are given $(a', b') \in A \times B$. Then $A \times \{b\}$ is connected, so $(a', b) \in U$. Similarly $\{a'\} \times B$ is connected, so $(a', b') \in U$. Thus $V = \emptyset$. ∎

It follows that any finite product of connected sets is connected.

**Infinite products.** Exercise: Show that an infinite product of connected sets is connected.

**Components.** Because fans are connected, every point $x \in X$ lies in a unique maximal connected set $A(x)$. This is called the *component* of $X$ containing $x$. Now it is easy to show that the closure of a connected set is connected, so $A(x)$ is closed (else it would not be maximal).

In general $A(x)$ is not open; for example, in the Cantor set $A(x) = \{x\}$ for all $x$. The components *are* open when $X$ has only finitely many components.

# 5 Compact spaces

In advanced calculus one learns that the interval $[a, b]$, and more generally any closed, bounded subset of $\mathbb{R}^n$, is compact.

Compact spaces $K$ have many important properties. For example, any continuous map $f : K \to \mathbb{R}$ achieves its maximum. Compact spaces are also 'bounded' in a strong sense, and thus they capture some of metric idea of a space of bounded diameter.

Unlike the definition of connectedness, the notion of compactness is subtle and requires some practice to use. It is almost always better to work with compact spaces that are *also Hausdorff*, and indeed Bourbaki has *redefined* the word 'compact' to include this condition.

**Finite covers.** Let $X$ be a topological space. A collection of open sets $(U_i)$ with $\bigcup U_i = X$ is an *open cover* of $X$. If we can extract from this covering a finite collection of open sets

$$(V_1, \ldots, V_k) = (U_{i_1}, \ldots, U_{i_k})$$

that still cover $X$, we say $(U_i)$ has a *finite subcover.* The space $X$ is compact if *every open cover of $X$ has a finite subcover.*

A more succinct statement is: if $\mathcal{U}$ is a collection of open sets and $X = \bigcup \mathcal{U}$, then there is a finite set $\mathcal{V} \subset \mathcal{U}$ such that $X = \bigcup \mathcal{V}$.

**Examples.** The real numbers are not compact, since the covering $\mathbb{R} = \bigcup_{\mathbb{Z}}(n, n+2)$ has no finite subcover. The interval $(0, 1]$ is noncompact as well: $\bigcup (1/n, 1]$ has no finite subcover. But $[0, 1]$ is compact. For example, if we write

$$[0, 1] = [0, \epsilon) \cup \bigcup_n (1/n, 1],$$

then $1/N < \epsilon$ for some $N$, and hence we have a finite subcover (by just two elements!) $[0, 1] = [0, \epsilon) \cup (1/N, 1]$.

**Two main results.** As for connectedness, there are 2 basic results, one abstract, one concrete.

**Theorem 5.1** *The continuous image of a compact space is compact.*

**Proof.** Suppose $A$ is compact and $f : A \to X$ is continuous. Let $B = f(A)$ and let $\bigcup U_i$ be an open cover of $B$. Then $\bigcup f^{-1}(U_i)$ is an open cover of $A$. Since $A$ is compact, the covering $f^{-1}(U_i)$ has a finite subcover, say $(f^{-1}(U_j) : j \in J)$, $|J| < \infty$. Since $U_j = f(f^{-1}(U_j))$, the sets $(U_j \,:\, j \in J)$ cover $B$. ■

In analysis one frequently discusses compactness in terms of sequences, but this does not suffice for general topological spaces. Thus it is interesting to prove compactness on the real line directly using the covering condition.

**Theorem 5.2** *The interval* $[0, 1]$ *is compact.*

**Proof.** Let $(U_i)$ be an open cover of $[0, 1]$. Let

$$A = \{x \in [0, 1] \,:\, \text{there is a finite subcover of } [0, x]\}.$$

Clearly $0 \in A$. We will show $A$ is open and closed. Then, since $[0, 1]$ is connected, we have will have $1 \in A$, which is the Theorem. In fact it is obvious that $A$ is open, since any cover that works for $[0, x]$ has $x \in U_i$ for some $i$; hence $B(x, r) \subset U_i$ for some $r > 0$, and hence the same cover works for $[0, x + r/2]$. The proof that $A$ is closed is similar. Suppose $x$ is a limit point of $A$. Of course $x \in B(x, r) \subset U_i$ for some $i$. Moreover $[0, x - r/2]$ admits a finite cover. Now add $U_i$ to this cover, and we have a finite cover of $[0, x]$. ■

Of course the same result holds for any other interval. We will shortly see the following basic result:

**Theorem 5.3** *A set* $K \subset \mathbb{R}^n$ *is compact iff it is closed and bounded.*

**Compact Hausdorff spaces.** We now explain a few points about compact sets versus closed sets.

**Theorem 5.4** *If* $X$ *is compact, then any closed set* $A \subset X$ *is also compact.*

**Proof.** Given any open cover of $A$, we can use $U = X - A$ to extend it to any open cover of $X$. We then pass to a finite cover, discard $U$, and obtain a finite open cover of $A$. ■

Here is a subtle point, whose proof is a beautiful application of compactness.

**Theorem 5.5** *If $X$ is a Hausdorff space, then any compact set $K \subset X$ is closed.*

**Proof.** Suppose $x \in X - K$. By the Hausdorff condition, for every $y \in K$ there is are disjoint neighborhoods $U_y$ and $V_y$ of $x$ and $y$ respectively. Then $K$ is covered by $\bigcup V_y$. Pass to a finite subcover, say $V_1, \ldots, V_k$ with corresponding neighborhood $U_1, \ldots, U_k$ of $x$.

> *Since a finite intersection of open sets is open, $U = U_1 \cap \cdots U_k$ is a neighborhood of $x$ as well.*

Now this neighborhood is disjoint from $K \subset V_1 \cup \cdots V_k$ and hence $x$ is not a limit point of $K$. Thus $K$ is already closed. ■

**Cofinite revisited.** Let $X$ be an infinite set (say $\mathbb{R}$) with the cofinite topology. Then $X$ is compact — as is any subset of $X$. But there are lots of subsets of $X$ that are not closed (any infinite set $A \neq X$).

**Continuity and compactness.** One great feature of compact spaces is that proofs that maps are homeomorphisms are twice as easy.

**Corollary 5.6** *A continuous bijection $f : X \to Y$ between compact Hausdorff spaces is a homeomorphism.*

**Proof.** It suffices to show that whenever $A \subset X$ is closed, so is $f(A)$. But if $A$ is closed, it is compact; thus $f(A)$ is also compact; and since $Y$ is Hausdorff, this means $f(A)$ is closed as well. ■

This result is certainly false without compactness. The simplest counterexample is the natural map $f : [0, 1) \to S^1 = \mathbb{R}/\mathbb{Z}$.

**Theorem 5.7** *A product of compact sets is compact.*

**Proof.** An open covering of $A \times B$ can be *refined* so its elements are all of the form $U \times V$, with $U$ and $V$ open. It suffices to extract a finite cover from this refinement.

Now for any $a \in A$, the set $\{a\} \times B$ is compact, so it has a finite cover of the form $\bigcup_1^n U_i \times V_i$. We may assume $a \in U_i$ for all $i$. Then $W = \bigcap U_i$ is a neighborhood of $a$, and $W \times B$ admits a finite cover. Since $A$ is compact, can cover it by finitely many open sets $W_i$ of this form. Each product $W_i \times B$ admits a finite cover, so their union $A \times B$ admits a finite cover as well. ■

**Corollary 5.8** *A set $F \subset \mathbb{R}^n$ is compact iff it is closed and bounded.*

A stronger result is due to Tychonoff:

**Theorem 5.9** *Given any collection of compact spaces $X_i$, the space $\prod X_i$ is compact in the product topology.*

Here is a curious remark: if the $X_i$ are nonempty, how do we even know that $\prod X_i$ is nonempty? To show this in general requires the Axiom of Choice!

We will prove Tychonoff's theorems for countable products of metric spaces later. The general case can be handled similarly, using nets and transfinite induction, although modern proofs are much more slick.

**Compactness and closed sets.** Let $F_n \subset \mathbb{R}$ be a sequence of nonempty closed sets with $F_1 \supset F_2 \supset \cdots$. Can it be that $\bigcap F_n$ is empty, even though at every finite stage it is nonempty? The answer is yes: e.g. take $F_n = [n, \infty)$.

But the answer is no for $F_n \subset [0, 1]$. More generally we have:

**Theorem 5.10** *Suppose $X$ is compact. Let $\mathcal{F} \subset \mathcal{P}(X)$ be a collection of closed sets such that $F_1 \cap \cdots F_k \neq \emptyset$ for all finite subsets $\{F_1, \ldots, F_k\} \subset \mathcal{F}$. Then $\bigcap \mathcal{F} \neq \emptyset$.*

**Proof.** If $\bigcap \mathcal{F} = \emptyset$ then $\bigcup \mathcal{U} = X$, where $\mathcal{U} = \{X - F \ : \ F \in \mathcal{F}\}$. Then $\mathcal{U}$ has a finite subcover, with $U_1 \cup \cdots \cup U_k = X$. Taking complements, we obtain elements of $\mathcal{F}$ such that $F_1 \cap \cdots \cap F_k = \emptyset$. ■

**Example.** From the ratios $f_{n+1}/f_n$ of the Fibonacci numbers

$$(1, 1, 2, 3, 5, 8, 13, 21, \ldots)$$

we obtain a nested sequence of intervals

$$(F_n) = ([1, 2], [3/2, 2], [3/2, 5/3], [8/5, 5/3], \ldots).$$

We have $\bigcap F_n = \{(1+\sqrt{5})/2\}$.

**Proof of Theorem 5.3.** If $K \subset \mathbb{R}^n$, then it is closed because $\mathbb{R}^n$ is Hausdorff, and it is bounded because $\bigcup B(0,n)$ has a finite subcover.

Conversely, if $K$ is bounded then it is contained in a product of intervals $[-M,M]^n$, which is compact. Thus if $K$ is also closed, it is a closed subset of a compact space, and hence compact. ■

**Calculus.** The following typical applications are crucial in science, economics, etc.

**Corollary 5.11** *A continuous function $f : X \to \mathbb{R}$ on a compact space achieves its maximum.*

**Proof.** In this case $f(X) \subset \mathbb{R}$ is compact, so it is closed and bounded. Thus $m = \sup f(X) < \infty$ and $m \in f(X)$, so $m = f(x)$ for some $x \in X$. ■

**Corollary 5.12** *Let $f : K \to \mathbb{R}$ be a smooth function on a closed, bounded domain in $\mathbb{R}^n$. Then $f$ achieves its maximum at a point $x \in \partial K$, or at a point $x \in \operatorname{int}(K)$ where $Df = 0$.*

**Nets*.** Sequences have a beautiful generalization called *nets* that restores their usefulness for general topological spaces.

A *directed set* $I$ is a partially ordered set such that for any $i, j \in I$, there is a $k \in I$ with $i < k$ and $j < k$. A *net* is a map $x : I \to X$ from a directed set into a topological space. We say $x_i \to y$ if for every open neighborhood $U$ of $y$, there is an $n \in I$ such that $x_i \in U$ for all $i > n$.

**Theorem 5.13** *If $x$ is a limit point of $A \subset X$, then there is a net $a_i \in A$ that converges to $x$.*

**Proof.** Let $I$ be the set of all neighborhoods of $A$, with $U < V$ if $V \subset U$. Since $x$ is a limit point of $A$, by the Axiom of Choice there is a net $a : I \to A$ such that $a_U \in U \cap A$. It is then a tautology that $a_U \to x$. ■

A subset $E \subset I$ is *cofinal* if for every $i \in I$ there is a $j \in E$ with $j > i$. If $J$ is another directed set, and $\pi : J \to I$ is an order–preserving map such that $\pi(J)$ is cofinal in $I$, we say $x_{\pi(j)}$ is a *subnet* of $x_i$.

**Theorem 5.14** *A space $X$ is compact iff every net $(x_i)$ in $X$ has a convergent subnet.*

**Proof.** Suppose every net has a convergent subnet. Let $\mathcal{F}$ be a family of closed sets with the finite intersection property. Choose a point $x_{\mathcal{A}} \in \bigcap \mathcal{A}$ for each finite subset $\mathcal{A} \subset \mathcal{F}$. Pass to a subnet so $x_i \to x \in X$. Now for any particular $F \in \mathcal{A}$, we have $F \in \mathcal{A}$ for all $\mathcal{A} > \{F\}$. Thus $x_i \in F$ for all $i$ 'sufficiently large', and hence $x \in F$. It follows that $x \in \bigcap \mathcal{F}$ and hence $X$ is compact.

The converse is similar. ∎

With this terminology in place, the product theorem for a pair of general compact spaces can be proved by passing twice to subnets as in the usual argument for metric spaces.

**Local and global perspective.** Here are a few formal similarities between the properties of compactness and connectedness:

- Both are preserved by continuous maps, $A \mapsto f(A)$;
- Both are preserved by products, $A \times B$; and
- The interval $[0, 1]$ is both compact and connected.

However compactness has the stronger *hereditary* property that if $X$ is compact and $A \subset X$ is closed, then $A$ is also compact. This means that once the ambient space $X$ is compact, compactness becomes a *local property*, while connectedness remains a *global property*.

# 6 Metric spaces

In this section we discuss the topology of metric spaces. Especially we show the metric space $C(K)$ is complete when $K$ is compact, and discuss *its* compact subsets (these are spaces of functions).

**Sequences.** Let $(X, d)$ be a metric space, given its usual topology: a basis for which consists of all the open balls $B(x, r)$ in $X$. Note that $X$ is *Hausdorff*, so limits are unique. Recall that $x_i \to x$ iff $d(x_i, x) \to 0$.

One of the main differences between $X$ and a general topological space is that we can think purely in terms of sequences. Thus:

1. The closure of $A \subset X$ is given by taking all limits of convergence sequences with $x_i \in A$.

2. A function $f : X \to Y$ between metric space is continuous iff $f(x_i) \to f(x)$ whenever $x_i \to x$.

3. The space $X$ is compact iff every sequence has a convergent subsequence.

**Completeness.** A *Cauchy sequence* in a metric space is a sequence $x_i \in X$ such that

$$\lim_{N\to\infty} \sup_{i,j>N} d(x_i, x_j) = 0.$$

We say $X$ is *complete* if every Cauchy sequences converges. This means whenever $d(x_i, x_j) \to 0$, we have $x_i \to x$ for some $x \in X$.

A Cauchy sequence is like a swarm of bees. In a complete space, you can see what they are after. Any metric space can be naturally embedded in a complete metric space. The real numbers $\mathbb{R}$ are the completion of $\mathbb{Q}$; this is their basic *raison d'être*.

(Paradox: it is tempting to *define* $\mathbb{R}$ as the completion of $\mathbb{Q}$, but then we need to first define complete metric spaces — without reference to $\mathbb{R}$!)

One can also complete the rational numbers with respect to the $m$-adic valuation, where $|m^k q| = m^{-k}$ for integers with $\gcd(m, q) < m$. For example, $\mathbb{Z}_{10}$ consists of decimal numbers which are infinite to the left. It is homeomorphic to $X^{\mathbb{N}}$, where $X = \{0, 1, 2, \ldots, 9\}$. For $p$ prime, the completion $\mathbb{Q}_p$ is a field, but $\mathbb{Q}_{10}$ is not. If $5^n$ accumulates on $x$ and $2^n$ accumulates on $y$, then $|x|_{10} = |y|_{10} = 1$ but $xy = 0$.

More analytical examples of a complete metric spaces come from measurable functions, such as $L^2[0, 1]$. Indeed, a major program in analysis is to study completions of spaces of functions with respect to various norms or metrics. The space $L^2[0, 1]$ is the completion of $C[0, 1]$ with respect to the norm

$$\|f - g\|^2 = \int_0^1 |f(x) - g(x)|^2 \, dx.$$

The theory of Lebesgue measurable functions is design to give concrete representatives for elements of this space, just as the real numbers describe elements of the completion of $\mathbb{Q}$.

**Sequential compactness.** Compactness has an alternative intuitive definition in metric spaces. We say $X$ is *sequentially compact* if every sequence has a convergent subsequence.

**Theorem 6.1** *For metric spaces, compactness and sequential compactness are equivalent.*

For the proof is it useful to first observe:

**Theorem 6.2** *A sequentially compact metric space $X$ has a countable base.*

**Proof.** It suffices to show that for each $n$, we can find a *finite* set of open balls $\mathcal{B}_n$ of radius $1/n$ that cover $X$. To see this, choose a maximal $E \subset X$ with $d(e, e') \geq 1/n$ for all $e \neq e'$ in $E$. Then $E$ is finite, since an infinite sequence of distinct points in $E$ would have no limit point, violating sequential compactness. We claim the balls $B(x, 1/n)$ with $x$ in $E$ cover $X$. Indeed, if some point $y \in X$ were left over, then we would have $d(y, x) \geq 1/n$ for all $x \in E$, so we could add $y$ to $E$.

It is now easy to check that $\mathcal{B} = \bigcup \mathcal{B}_n$ gives a countable base for $X$. ∎

**Proof of Theorem 6.1.** Suppose $X$ is compact and $x_n$ is a sequence in $X$. We may assume the elements of the sequence are all distinct. Let $F_n$ be the closure of the set $\{x_i : i \geq n\}$. Then $E = \bigcap F_n$ is nonempty by compactness. If $y \in E$ than for any $n > 0$ we can find an $x_{i_n} \in B(y, 1/n)$. Thus we have a subsequence converging to $y$.

The converse is a little trickier. Suppose $X$ is sequentially compact, and $\mathcal{U}$ is an open cover of $X$. As we have seen, $X$ has a countable base. Thus we can assume $\mathcal{U}$ is a *countable* cover of $X$, by choosing for each basis element $B$ a single element of $\mathcal{U}$ that contains it (if one exists).

We can now order the open sets as $U_1, U_2, \ldots$, and we want to show that $V_n = \bigcup_1^n U_i$ is equal to $X$ for some $n$. Equivalently, setting $F_n = X - V_n$, it suffices to show that if we have a sequence of nonempty closed sets $F_1 \supset F_2 \supset \cdots$, then $F = \bigcap F_i \neq \emptyset$. This is easy: just choose $x_i \in X$, and use sequential compactness to obtain a subsequence converging to $y$; then $y \in F$. ∎

**Corollary 6.3** *A compact metric space is complete.*

**Proof.** Any Cauchy sequence $x_i$ has a subsequence with a limit $y$; but then $x_i \to y$ as well. ∎

**Lebesgue number and uniform continuity.** Here are 2 useful facts about functions and coverings in metric spaces.

**Theorem 6.4** *Let $\mathcal{U}$ be an open cover of a compact metric space $K$. Then there exists an $r > 0$ such that for any $x \in K$, we can find $U \in \mathcal{U}$ such that $B(x,r) \subset U$.*

**Proof.** Let $V_s \subset K$ be the set of $x$ such that $B(x,s)$ is contained in some element of $\mathcal{U}$. Clearly $\bigcup_{s>0} V_s = K$. Since $K$ is compact, we have $K = V_r$ for some $r > 0$. and then $r = \min(s_1, \ldots, s_i)$ works. ∎

**Corollary 6.5** *Let $f : K \to \mathbb{R}$ be continuous. Then for any $\epsilon > 0$ there exists an $r > 0$ such that $d(x,y) < r \implies |f(x) - f(y)| < \epsilon$.*

**Proof.** Let $\mathcal{U}$ be the covering of $K$ by the preimages under $f$ of all intervals of length $\epsilon$ in $\mathbb{R}$, and apply the preceding result. ∎

**Uniform limits.** Let $Y^X$ denote the set of all functions $f : X \to Y$, and let $C(X,Y) \subset Y^X$ denote the subset of continuous functions.

Suppose $Y$ is a metric space. Then we can define a distance between functions by

$$d(f,g) = \sup_X |f(x) - g(x)|.$$

This is not quite a metric, since it may be infinite. We can rectify this by simply setting $d(f,g) = 1$ if the supremum above is $> 1$. We say $f_n \to g$ *uniformly* if $d(f_n, g) \to 0$. We call $d$ the uniform metric.

**Theorem 6.6** *If $Y$ is complete, then so is $(Y^X, d)$.*

**Proof.** Suppose $f_n \in Y^X$ is a Cauchy sequence. Then so is $f_n(x)$; thus $f_n(x) \to g(x)$, and

$$d(f_n, g) \leq \sup_{m>n} d(f_n, f_m) \to 0$$

since $(f_n)$ is a Cauchy sequence. ∎

**Theorem 6.7** *A uniform limit of continuous functions is continuous. In other words, $C(X,Y)$ is closed in $Y^X$.*

**Proof.** Suppose $f_n \in C(X,Y)$ converge uniformly to $g : X \to Y$. Let $V \subset Y$ be open. We must show that $U = g^{-1}(V)$ is open. Suppose $x \in g^{-1}(V)$, and $y = g(x) \in V$. Then $B(y,r) \subset V$ for some $r > 0$.

Let $U_n = f_n^{-1}(B(y,r/2))$. Clearly $U_n$ is open for all $n$. Choose $N$ so $d(f_n,g) < r/2$ for all $n > N$. Then if $d(f_n(x),y) < r/2$, we have $d(g(x),y) < r$, and hence

$$U_n = f_n^{-1}(B(y,r/2)) \subset g^{-1}(B(y,r)) \subset U$$

for all $n > N$. Moreover $f_n(x) \to g(x) = y$, so $x \in U_n$ for all $n \gg 0$. It follows that $x \in U_n \subset U$ for all $n \gg 0$, and thus $U$ is open. ∎

**Corollary 6.8** *If $Y$ is complete, then so is $C(X,Y)$.*

**Computer programs.** Suppose you are asked to write a computer program that takes as input $x \in [0,1]$ and outputs $F(x) = 1$ if $x = 1$ and $F(x) = 0$ otherwise.

Unfortunately you have not yet learned how to use branching in your program. So you propose a program that just uses iteration: it computes $f_n(x) = x^n$, and outputs $F(x) = \lim f_n(x)$.

Now of course, in practice you can only compute finitely many steps. And even if $x < 1$, you won't get $f_n(x) < 1/2$ until you have iterated about $n = |1/\log_2(x)|$ times. More and more iterations are required, the closer $x$ is to one! This *nonuniformity* results from the fact that you are trying to compute a *discontinuous* function $F(x)$, without using if statements. So even though $f_n \to F$ pointwise, it *cannot* do so uniformly — otherwise $F$ would be continuous.

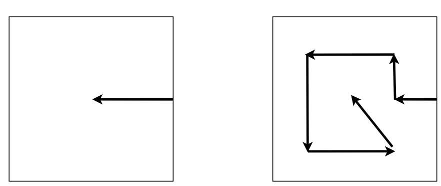


Figure 2. Peano curve motif.

**Peano curves.** An amusing application of completeness is the construction of a *Peano curve*, that is a continuous, surjective map $f : I \to I^2$. This shows that the dimension of a set can go up when we take its continuous image!

Cantor was already amazed when he saw that there is a set-theoretic bijection between $I$ and $I^2$. (The basic idea of this is easy: send $0.x_1x_2x_3\ldots$ to $(0.x_1x_3x_5\ldots, 0.x_2x_4x_6\ldots)$.) More interesting is the fact that a surjective map $I \to I^2$ can be made continuous.

Exercise: there is no continuous *bijection* $f : I \to I^2$. (Use connectivity.)

To construct a Peano curve it suffices to show:

**Theorem 6.9** *There exists a uniformly convergent sequence of functions $f_n : I \to I^2$ such that $f_n(I)$ enters every dyadic square of side $2^{-n}$.*

One such construction (there are many!) is shown in Figure 2. Whenever we have a segment $J$ such that one endpoint of $f(J)$ passes through the center of a dyadic square $S$, we can modify $f|J$ so that it also passes through the centers of the 4 adjacent subsquares. In doing so we never move $f(x)$ a distance greater than $\operatorname{diam}(S) = 2^{-n}$. Since $\sum 2^{-n}$ is finite, the sequence converges uniformly. The sequence is shown in Figure 3.

**Products of metric spaces.** Using sequential compactness, it is quite easy to see that the product of two compact metric spaces is compact. Namely, if $(x_i, y_i)$ is a sequence in $X \times Y$, we can pass first to a subsequence such that $x_i \to x$, and then to a subsequence such that $y_i \to y$. Then $(x_i, y_i) \to (x, y)$.

By diagonalization, it is similarly easy to see that a countable product of compact metric spaces, $\prod_1^\infty X_i$, is compact.

We can now give a 'new' proof that $[0, 1]$ is compact: the product $2^{\mathbb{N}}$ is obviously compact, and there is a surjective, continuous function $f : 2^{\mathbb{N}} \to [0, 1]$ given by $f(a_0, a_1, a_2, \ldots) = \sum 2^{-i-1} a_i$.

**Characterization of compact metric spaces.** As we have seen, $K \subset \mathbb{R}^n$ is compact iff it is closed and bounded. Here is a generalization. We say a metric space $X$ is *totally bounded* if for every $r > 0$, there is a covering of $X$ by finitely many balls of radius $r$.

**Theorem 6.10** *A metric space is compact iff it is complete and totally bounded.*

**Proof.** We have already seen that a compact metric space is complete; and total boundedness follows by taking a finite subcover of the covering by all balls of radius $r$.

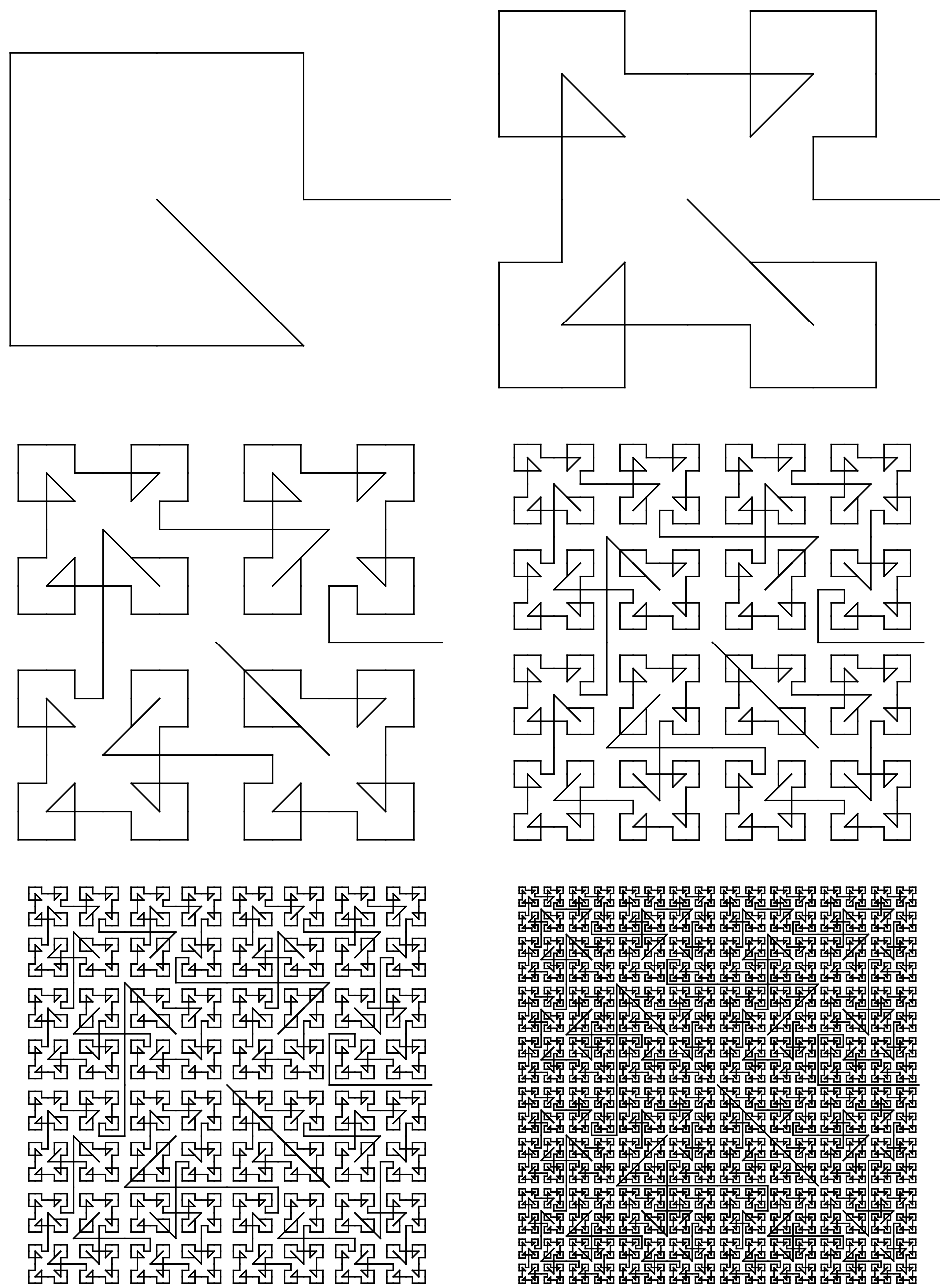


Figure 3. Approximate Peano curves.

Now suppose $X$ is complete and totally bounded. Let $x_n$ be a sequence in $X$. Given $r > 0$, we can cover $X$ by finitely many $r$ balls. Then there is a subsequence that lies entirely in one of these balls; it satisfies $d(x_i, x_j) < 2r$ for all $i, j$. Diagonalizing, we obtain a subsequence of $(x_n)$ which is also a Cauchy sequence. By completeness, this subsequence has a limit $y$, and hence $X$ is (sequentially) compact. ∎

**Functions on a compact space.** Whenever $K$ is compact, we can make $C(K) = C(K, \mathbb{R})$ into a complete metric space by setting

$$d(f, g) = \|f - g\| = \sup_X |f(x) - g(x)|.$$

This supremum is finite because $X$ is compact.

**The Arzela–Ascoli theorem.** As a vector space, $C[0, 1]$ is infinite–dimensional. Unlike $\mathbb{R}^n$, its closed, bounded subsets are not compact. For example, $f_n(x) = x^n$ has no convergent subsequence; neither does $f_n(x) = \sin(nx)$.

One of goals is to describe the compact subsets of $C[0, 1]$ and more generally of $C(X)$, when $X$ is a metric space. Here is a typical application of the preceding fact.

**Theorem 6.11** *Let $F \subset C([0, 1])$ be the set of differentiable functions with $|f(x)| \le 1$ and $|f'(x)| \le 1$. Then $\overline{F}$ is compact.*

In other words, a sequence of bounded functions with bounded derivatives has a uniformly convergent subsequence.

**Proof.** To understand this result, it is important to first see that the unit ball in $C([0, 1])$, unlike the unit ball in $\mathbb{R}^n$, is *not* compact. (Consider the sequence $f_n(x) = x^n$.) But in this example, $f_n'(1) = n \to \infty$. We cannot make an example like this if we keep $|f'|$ bounded!

The reason for this is that $F$ (or equivalently $\overline{F}$) is totally bounded. Here is a proof. Let us say $f, g \in C([0, 1])$ are close along the $1/n$ grid if

$$|f(a/n) - g(a/n)| < 1/n$$

for $a = 0, 1, \ldots, n$. It is easy to find a finite set $G_n \subset F$ such that every $f \in F$ is close to some $g_n \in G_n$ along the $1/n$ grid. But then, by the intermediate value theorem, we also have

$$|f(x) - g_n(x)| \le 2/n$$

for all $x$. That is, the balls $B(g_n, 2/n)$ cover $F$ and the balls $B(g_n, 3/n)$ cover $\overline{F}$. *Thus $\overline{F}$ is totally bounded.*

Since $C(K)$ is complete, so is $\overline{F}$. Thus $\overline{F}$ is compact. ∎

**Equicontinuity.** More generally, if $K$ is a compact metric space and $F \subset C(K)$ is a family of continuous functions on $K$, we say $F$ is *equicontinuous* if for all $\epsilon > 0$ there exists a $\delta > 0$ such that

$$d(x, y) < \delta \implies |f(x) - f(y)| < \epsilon$$

for all $f \in F$. (In the example above, we can take $\delta = \epsilon$.) This property allows us to replace 'close on a grid' with 'close throughout $K$' as we did in the proof above. Thus the same argument shows:

**Theorem 6.12 (Arzela–Ascoli)** *Let $K$ be a compact metric space, and let $F \subset C(K)$ be an equicontinuous family of uniformly bounded functions. Then $\overline{F}$ is compact.*

**Complex analysis and normal families.** This result is not so powerful in calculus but it is very powerful in complex analysis. For example, one can use it to show that any sequence of analytic functions

$$f_n : U \to \mathbb{C},$$

satisfying $|f_n(z)| \leq M$ for all $n$ and $z \in U$, has a subsequence that converges uniformly on compact subsets of $U$ to another *analytic* function $g(z)$. This is because, by Cauchy's integral formula, we have

$$|f'(z)| = \left| \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{(z - \zeta)^2} \, d\zeta \right| \leq \frac{M}{d(z, \partial U)}.$$

**Baire category.** Here is a final remark about complete metric spaces. If $X$ is compact, then any nested intersection $\bigcap F_i$ of nonempty closed sets is nonempty.

**Theorem 6.13 (Baire category)** *If $X$ is complete, then the intersection $\bigcap_1^\infty U_i$ of any sequence of dense open sets is nonempty (and dense).*

**Proof.** By density we can choose a nested sequences of closed balls $B(x_i, r_i) \subset U_i$ with $r_i \to 0$. Then $(x_i)$ forms a Cauchy sequence, and $y = \lim x_i$ lies in $B(x_i, r_i)$ so it lies in $\bigcap U_i$. This shows $\bigcap U_i$ is nonempty; and by varying the choice of $B(x_1, r_1)$, we see it is dense. ∎

**Corollary 6.14** *If $X$ is a complete metric space and $F_1 \subset F_2 \subset \cdots$ is an increasing sequence of closed sets with $X = \bigcup F_i$, then one of them has nonempty interior.*

**Example.** (A challenging homework problem.) Find a sequence of open sets $U_i \subset \mathbb{R}$ such that $\bigcap U_i = \mathbb{Q}$.

Answer: this is impossible! Let $\mathbb{Q} = \{p_1, p_2, \ldots\}$ and let $V_i = U_i - \{p_1, \ldots, p_i\}$. Since $U_i$ contains $\mathbb{Q}$, each $V_i$ is a dense open set; but if $\bigcap U_i = \mathbb{Q}$ then $\bigcap V_i = \emptyset$.

Baire category is frequently useful in the construction of counterexamples. For example, rather than showing there exists a nowhere differentiable function $f \in C[0,1]$, one can show that a *generic* function is nowhere differentiable.

**Example: Liouville numbers.** We say $x \in \mathbb{R}$ is a Liouville number if $x$ is irrational and, for any $k$, there are infinitely many rational numbers with

$$\left| x - \frac{p}{q} \right| \le \frac{1}{q^k} \cdot$$

In other words, $x$ is very well approximated by rational numbers. For example, $x = \sum_0^\infty 1/10^{n!}$ is a Liouville number, since $x_m = \sum_0^m 1/10^{n!} = p_m/q_m$ with $q_m = 10^{m!}$, and

$$|x - p_m/q_m| \le 2 \cdot 10^{-(m+1)!} \le 10^{-km!} = 1/q_m^k$$

for all $m$ large enough.

If we set $U_k = \bigcup B(p/q, 1/q^k)$ then the set of Liouville numbers if given by $\bigcap_1^\infty U_k - ratls$, and thus:

**Theorem 6.15** *A generic real number is a Liouville number.*

On the other hand, it is not hard to show that a number chosen at random in $[0,1]$ is *not* a Liouville number.

# 7 Normal spaces

There are many separation axioms that interpolate between the Hausdorff spaces and metrizable spaces. In this section we will focus on the most important such condition, *normality.*

A topological space is said to be *normal* if (a) points are closed and (b) any pair of disjoint closed sets can be engulfed by a pair of disjoint open sets.

More formally, if $A$ and $B$ are closed and $A \cap B = \emptyset$, then there exist open set $U$ and $V$ with $A \subset U$, $B \subset V$ and $U \cap V = \emptyset$. Note that a normal space is Hausdorff (take $A$ and $B$ to be single points, which are closed as part of the definition of normality).

Any metric space is normal: we can simply let

$$U = \{x \,:\, d(x, A) < d(x, B)\} \quad \text{and} \quad V = \{x \,:\, d(x, B) < d(x, A)\}.$$

Slightly more subtle is:

**Theorem 7.1** *A compact Hausdorff space is normal.*

**Proof.** We first prove $X$ is regular: a closed set $A$ and a point $b \not\in A$ have disjoint open neighborhoods.

By the Hausdorff assumption, for each $a \in A$ we have disjoint neighborhoods $U_a$ of $a$ and $V_a$ of $b$. Finitely many of these cover $A$; call their union $U$, and let $V_b$ be the intersection of the correspond open sets $V_b$. Then $U$ and $V$ provide disjoint neighborhood of $A$ and $b$.

We can now prove $X$ is normal by the same style of argument: by regularity, given disjoint closed sets $A$ and $B$, for each $a \in A$ we can find disjoint neighborhoods $U_a$ of $a$ and $V_b$ of $B$. The union $U$ of finitely many of these $U_a$ covers $A$, and the intersection $V$ of the corresponding open sets $V_a$ contains $B$. ■

Here are two general results we will prove about normal spaces. They are often called Urysohn's Lemma and Urysohn's Metrization Theorem.

**Theorem 7.2** *Let $A$ and $B$ be disjoint closed subsets of a normal space $X$. Then there exists a function $f \in C(X)$ such that $f|A = 0$ and $f|B = 1$.*

**Theorem 7.3** *A normal space with a countable base is metrizable.*

**Proof of Theorem 7.2.** Using normality it is easy to inductively construct a sequence of closed sets $A_s$ indexed by the numbers of the form $s = p/2^n$, $0 \leq p/2^n \leq 1$, such that

1. $A_0 = A$, $A_1 = X$,

2. We have $A_s \subset \text{int}(A_t)$ whenever $s < t$.

3. We have $A_s \subset X - B$ whenever $s < 1$.

Once this is done, we define $f : X \to [0, 1]$ by

$$f(x) = \inf\{s \,:\, x \in A_s\}.$$

From the conditions above we get $f|A = 0$, $f|B = 1$. For continuity, first note that

$$f^{-1}((-\infty, a)) = \{x \,:\, f(x) < a\} = \bigcup_{s<a} A_s = \bigcup_{s<a} \text{int}(A_s)$$

is a union of open sets, hence open. Similarly

$$f^{-1}((b, +\infty)) = \{x \,:\, f(x) > a\} = \bigcup_{s>b} (X - A_s)$$

is also open. Since these two types of intervals generate the topology on the real numbers, $f$ is continuous. ■

**Remark.** Even though $f|A = 0$, we might not have $f^{-1}(0) = A$, and similarly for $B$.

**Maximal ideals.** An important consequence of Urysohn's Lemma is that if $K$ is a compact Hausdorff space, then the functions in $C(K)$ separate the points of $K$. Consequently any homomorphism of algebras $\phi : C(K) \to \mathbb{R}$ comes from a point evaluation (exercise).

**Proof of Theorem 7.3.** First construct, for every pair of elements in the countable base $B, B'$ with $\overline{B} \subset B'$, a function $f : X \to [0, 1]$ which is 1 on $B$ and 0 outside $B'$. Let the countable set of such functions be $(f_1, f_2, \ldots)$, and let

$$\phi : X \to [0, 1]^{\mathbb{N}}$$

be the tautological embedding:

$$\phi(x) = (f_1(x), f_2(x), \ldots).$$

We give the target the product topology. Now the Hilbert cube $[0, 1]^{\mathbb{N}}$ is certainly metrizable. so to complete the proof it suffices to show that $\phi$ is an embedding.

It is clear that $\phi$ is continuous. It is also $1-1$, since if $x \neq y$ we can find $B, B'$ as above with $x \in B$ and $y \notin B'$. (Thus if $X$ were compact, we would be done.)

We must show that $\phi : X \to \phi(X)$ is an open map. For this, we must show that for any open set $U$ in $X$ and $\phi(x) \in \phi(U)$, there is a neighborhood $V$ of $\phi(x)$ in $[0,1]^{\mathbb{N}}$ such that

$$\phi(X) \cap V \subset \phi(U).$$

(This shows $\phi(U)$ is open in the subspace topology.) To do this, choose $B$ and $B'$ as above with $x \in B \subset B' \subset U$, and let $f_i$ be the function that is 1 on $B$ and 0 outside $B'$.

Let $V \subset [0,1]^{\mathbb{N}}$ be the open set defined by the condition $(f_i > 1/2)$. Then if $\phi(y) \in V$, we have $y \in B'$ and hence $\phi(y) \in \phi(U)$, which completes the proof. ■

**Note.** One can also metrize a regular space with a countable base.

**Compactifications.** It is often useful, given a topological space $X$ which is not compact, to try to find a *compactification* $\overline{X}$ of $X$. This means we find a compact space $\overline{X}$, and an inclusion $X \subset \overline{X}$ which is a homeomorphism to its image, such that $X$ is dense in $\overline{X}$.

The problem is similar to that of metric completion, except that the answer is not unique. For example, we can compactify $(0,1)$ as a circle or as a an interval $[0,1]$.

**Locally compact spaces.** Let $X$ be a Hausdorff space. If $X$ has a basis $\{B\}$ such that $\overline{B}$ is compact for all $B$, we say $X$ is *locally compact*. The *one-point compactification* $X^*$ of $X$ is defined by adding the single point $\infty$ to $X$, and declaring that any set of the form $(X-K) \cup \{\infty\}$, with $K$ compact, is a neighborhood of infinity.

For example, the 1-point compactification of $\mathbb{R}^2$ is the 2-sphere.

Exercise: show $X*$ is compact.

**Compactification by ends.** Another compactification is given by $\overline{X} = X \cup \epsilon(X)$, where $\epsilon(X)$ is the space of *ends* of $X$, defined in terms of the components of $X-K$. Rather than giving a formal definition we just mention a few examples: (i) $\overline{X} \cong [0,1]$ for $X = \mathbb{R}$; (ii) $\overline{X} \cong S^n \cong X^*$ for $X = \mathbb{R}^n$, $n > 1$; (iii) $\overline{X} \cong X \cup d \times (d-1)^{\mathbb{N}}$ when $X$ is a regular tree of degree $d$.

**The Stone-Čech compactification.** Now suppose $X$ is a normal space. The biggest compactification of $X$, denoted $\beta(X)$, is defined as follows. Let

$I(X)$ be the set of all continuous maps $f : X \to [0,1]$. Then $[0,1]^C$ is compact, by Tychonoff's theorem. Define

$$\iota : X \to [0,1]^{I(X)}$$

by $\iota(x)_f = f(x)$. Then $\beta(X)$ is, by definition, the closure of $\iota(X)$.

The proof that $\iota$ is a homeomorphism to its image is similar to the proof of the metrizability theorem above.

This compactification is biggest in the following technical sense: any continuous map $f : X \to K$, where $K$ is a compact Hausdorff space, extends to a continuous map $\beta(f) : \beta(X) \to K$.

**The Stone-Čech compactification of $\mathbb{N}$.** The case $X = \beta(\mathbb{N})$ is already exotic. Here $\mathbb{N}$ has the discrete topology, so *every* map $a : \mathbb{N} \to [0,1]$ is continuous. In other words, $C$ consists of all sequences with $0 \le a_n \le 1$, and every sequence has a continuous extension to $\beta(\mathbb{N})$.

Let $\lim_\xi a_n$ denote the value of $a$ at a point $\xi \in \beta(\mathbb{N})$ with $\xi \notin \mathbb{N}$. This 'generalized limit' is not only linear and defined for all bounded sequences, but it also satisfies

$$\lim_\xi (a_n b_n) = (\lim_\xi a_n)(\lim_\xi b_n).$$

It agrees with the usual limit when $a_n$ is convergent. In fact the points of $\beta(\mathbb{N})$ can be identified with the *ultrafilters* on $\mathbb{N}$.

# 8 Algebraic topology and homotopy theory

The second main topic of this course will be elementary algebraic topology. The idea of algebraic topology is to associate to a continuous topological space $X$ a discrete algebraic object $A(X)$, with the aim of telling spaces apart. For this to work, we want the map $X \mapsto A(X)$ to be a *functor*; that is, if $f : X \to Y$ is continuous, we should also get a map

$$f_* : A(X) \to A(Y)$$

that respects the algebraic structure of $A(X)$.

**Homotopy.** Since $A(X)$ is discrete, if we continuously deform $f$, we will not change $f_*$. This suggests that $A(X)$ and $A(Y)$ will be the same if $X$ is *homotopy equivalent* to $Y$. Here is a precise definition.

A pair of continuous maps $f, g : X \to Y$ are *homotopic* if there is a continuous map $F : X \times [0,1] \to Y$ such that $F_0 = f$ and $F_1 = g$.

**Example.** Let $S^1 = \mathbb{R}/\mathbb{Z}$. There is an intuitive notion of degree for a map $f : S^1 \to S^1$, with $\deg(f) = n$ when $f(x) = nx \bmod 1$. It turns out that $f$ and $g$ are homotopic iff they have the same degree. In particular, the identity map is not homotopic to a constant map. One of our goals will be to prove this.

**Homotopy equivalence.** A *homotopy equivalence* between spaces is a map $f : X \to Y$ such that there exists a map $g : Y \to X$ with $g \circ f$ and $f \circ g$ *homotopic* to the identity on $X$ and $Y$ respectively.

**Retraction and contractible spaces.** A special case arises when the inclusion $A \subset X$ is a *deformation retract* of $X$. This means we have a continuous family of maps $F_t : X \to X$ such that (i) $F_t(x) = x$ for all $x \in A$, (ii) $F_0(x) = x$, and (iii) $F_1(x) \in A$. The final map $F_1$ is called a *retraction* of $X$ onto $A$. By continuous, we mean $(t, x) \mapsto F_t(x)$ is a continuous function on $[0,1]^2$

For example, the space $X = \mathbb{R}^n$ retracts onto the origin $A = \{0\}$. In this case we say $X$ is *contractible* (homotopy equivalent to a point).

**Example:** A theta, a dumbbell and a figure 8 are all homotopy equivalent. The coolest proof of this is to slightly thicken these sets in the plane. All of the thickenings are homeomorphic, and all retract onto the original set.

The punctured plane $X = \mathbb{R}^2 - \{0\}$ retracts onto the unit circle $A = S^1 \subset X$. The retraction can be given in polar coordinates by $F_t(e^s, \theta) = (e^{(1-t)s}, \theta)$. It is just obtained by retracting each ray through the origin to its intersection with the circle.

**The punctured torus.** It is an interesting an initially unexpected fact that a punctured torus is homotopy equivalent to a bouquet of 2 circles. What can you say for a punctured surface of genus 2?

In fact $X = S^1 \times S^1 - B$ is homeomorphic to the rotary with an underpass shown in Figure 1. If we think of $S^1 \times S^1$ as a quotient of the square, then it is easy to see that $X$ retracts onto $S^1 \vee S^1$. But $X$ represents a *different* thickening of this region from the ones coming from a dumbbell or theta or figure eight in the plane. Namely $X$ has only one boundary component, while the other thickenings have three. This reflects the facet that $X$ has a handle — it does *not* embed in the plane.

# 9 Categories and paths

We begin a more detailed treatment with a quick review of categories and groups.

**Categories.** A category is a collection of *objects* $A, B, C$ and an association to each pair of objects a family of *morphisms* (or arrows), $\mathrm{Mor}(A, B)$. We are also given a composition law:

$$\mathrm{Mor}(A, B) \times \mathrm{Mor}(B, C) \to \mathrm{Mor}(A, C).$$

The main axiom is that composition is associative:

$$(f \circ g) \circ h = f \circ (g \circ h)$$

whenever both sides are defined. We also require that for every object $A$ we have a morphism $\mathrm{id}_A \in \mathrm{Mor}(A, A)$ which behaves as the identity. This means

$$f \circ \mathrm{id}_A = \mathrm{id}_B \circ f = f$$

for all $f \in \mathrm{Mor}(A, B)$. By considering

$$\mathrm{id}_A \circ \mathrm{id}'_A,$$

we see that $\mathrm{id}_A$ is unique.

Warning: often the collection of all objects of a category is too big to be a set, e.g. it might be the 'collection' of all sets. We will ignore this logical difficulty, which can be circumvented by considering a category as a 'class'.

Examples:

1. The category of all sets, with $\mathrm{Mor}(A, B)$ as all maps $f : A \to B$.
2. The category of sets, with $\mathrm{Mor}(A, B)$ as the set of relations (subsets $f \subset A \times B$).
3. The category of all finite–dimensional vector spaces over $\mathbb{R}$, with $\mathrm{Mor}(A, B)$ as the set of linear maps.
4. The category of groups, with $\mathrm{Mor}(A, B)$ as group homomorphisms.
5. The category of all topological spaces, with $\mathrm{Mor}(A, B)$ as the set of continuous maps.

6. The category of groups, with the morphisms as group isomorphisms.

7. A partially ordered set can be made into the objects of a category, where the morphisms all move upwards.

8. The vertices of a directed graph with form a category, with directed paths as morphisms.

9. A group is a category with one object, in which all morphisms are isomorphisms.

10. The matrices $\mathrm{M}_n(\mathbb{R})$ form the objects of a category, with $\mathrm{Mor}(A, B) = \{C : CA = B\}$. (Any ring with identity could play the role of $\mathrm{M}_n(\mathbb{R})$ here.)

11. The cobordism category. In the case of dimension two, the objects are finite unions of circles, and the morphisms are surfaces they bound.

Note: $\mathrm{Mor}(A, B)$ may be empty, but $\mathrm{Mor}(A, A)$ always contains $\mathrm{id}_A$.

**What is a map?** When a category is understood, a *map* generally refers to a morphism in the category. Thus in the sequel, where we work in the category of topological spaces, we will often discuss a map $f : X \to Y$; the assumption that $f$ is *continuous* is implicit.

**Isomorphisms.** We say $f \in \mathrm{Mor}(A, B)$ is an isomorphism if there exists a $g \in \mathrm{Mor}(B, A)$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$. The set of all isomorphisms from $A$ to itself is called the *automorphism group* $\mathrm{Aut}(A)$. By applying associativity to the expression

$$g' \circ f \circ g,$$

we find:

*The inverse of an isomorphism is unique.*

We let $\mathrm{Isom}(A, B) \subset \mathrm{Mor}(A, B)$ denote the set of all isomorphisms for $A$ to $B$, and we let $\mathrm{Aut}(A) = \mathrm{Isom}(A, A)$.

**Groups.** A *group* is a set $G$ equipped with a product operation $G \times G \to G$ denoted $f * g$, an inverse map $G \to G$, denoted $g \mapsto g^{-1}$, and an identity element denoted $e \in G$, such that:

1. $e * g = g * e = e$ for all $g \in G$,

2. $g * g^{-1} = g^{-1} * g = e$a for all $g \in G$, and

3. $(f * g) * h = f * (g * h)$ for all $f, g, h \in G$.

In any category, $\mathrm{Aut}(A)$ is a group. Conversely, a group is the same thing as a category with one object where all morphisms are isomorphisms.

Examples: if $A$ is a set with $n$ elements, then $\mathrm{Aut}(A) \cong S_n$. If $A$ is a vector space of dimension $n$ over $\mathbb{R}$, then $\mathrm{Aut}(A) \cong \mathrm{GL}_n(\mathbb{R})$.

**Proposition 9.1** *If $A \cong B$ then $\mathrm{Aut}(A) \cong \mathrm{Aut}(B)$.*

The second isomorphism in general *depends* on the choice of a map from $A$ to $B$.

**Groupoids.** A *groupoid* is a category in which every morphism is an isomorphism. Basic examples are: sets with bijections; vector spaces or groups, with isomorphisms; topological spaces with homeomorphisms.

**The category of paths.** (Also known as the fundamental groupoid.) Here is a very useful and perhaps surprising example of a groupoid. Let $X$ be a topological space. Consider the category $\mathcal{P}(X)$ whose objects are the *points* $A, B, C \ldots \in X$.

A *path* from $A$ to $B$ is a continuous map $f : [0,1] \to X$ with $f(0) = A$ and $f(1) = B$. Two paths $f_0$ and $f_1$ are *homotopic* if there is a continuous family of paths $f_s(t)$ from $A$ to $B$, $s \in [0,1]$, interpolating between them. This notion gives an equivalence relation on paths. We denote the homotopy class of a path by $[f]$, and let

$$\mathrm{Mor}(A, B) = \{[f] \ : \ f \text{ is a path from } A \text{ to } B\}.$$

The *identity* $id_A = [c_A] \in \mathrm{Mor}(A, A)$ is the class of the constant path, $c_A(t) = A$.

**Question.** When is $\mathrm{Mor}(A, B)$ empty?

The interesting part is composition. Once we have a path $f$ from $A$ to $B$ and a path $g$ from $B$ to $C$, we can form a new path by

$$(f * g)(t) = \begin{cases} f(2t) & \text{for } t \in [0, 1/2], \text{ and} \\ g(2t-1) & \text{for } t \in [1, 2, 1]. \end{cases}$$

Actually the map is given by

$$[f] * [g] = [f * g].$$

One must check that this map is well–defined. But deformations of $f$ and $g$ paste together to give deformations of $f * g$, so it is.

Figure 4 explains why $[c_A \circ f] = [f]$ and $[(f * g) * h] = [f * (g * h)]$.

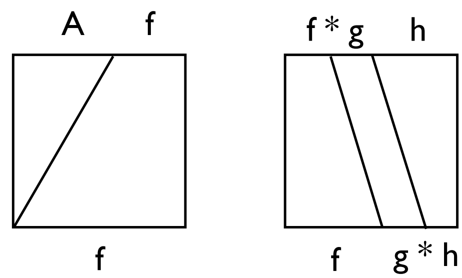



Figure 4. Proof that the category of paths really is a category.

**Inverses.** Finally we need to check that every path $f(t)$ has an inverse. But this is just that path defined by $g(t) = f(1 - t)$. Then $[g * f] \in \mathrm{Mor}(A, A)$ is the same as $[c_A]$, since $g * f$ can be deformed to a point, like slurping up a single noodle of spaghetti.

**The category of unparameterized paths.** We could also have identified paths $f$ and $g$ if $f(t) = g(h(t))$ for some homeomorphism $h : [0, 1] \to [0, 1]$ fixing the endpoints. The result would still be a category; everything would go through except the existence of inverses. The result is a continuous version of the category of paths in a graph.

**The invariant $\boldsymbol{\pi_0(X)}$.** As in any category, we can classify the objects up to isomorphism, i.e. introduce the equivalence relation $A \sim B$ if there exists an isomorphism from $A$ to $B$. For example, finite-dimensional vector spaces over $\mathbb{R}$ are classified up to isomorphism by their dimensions.

The *set* of isomorphism classes in the category of paths is called the space of *path components*: $\pi_0(X)$. Usually when dealing with the fundamental group, we assume our space is path–connected.

**The invariant $\boldsymbol{\pi_1(X)}$: the fundamental group.** The fundamental group of a space, in terms of the category of paths, is now defined by:

$$\pi_1(X, A) = \mathrm{Mor}(A, A) = \mathrm{Aut}(A).$$

It is simply the group of all loops based at $A$.

Exercise: a path $p$ from $A$ to $B$ determines an isomorphism

$$\pi_1(X, A) \cong \pi_1(X, B),$$

given by $f \mapsto p * f * p^{-1}$.

**Dependence on base point.** If $X$ is path connected, then for any two $A, B \in X$ we have $A \cong B$ in the category of paths, and thus $\mathrm{Aut}(A) \cong \mathrm{Aut}(B)$. This shows:

**Theorem 9.2** *If $X$ is path connected, then the group $\pi_1(X, x)$ is independent of $x$ up to isomorphism. More precisely, any path from $x$ to $y$ determines an isomorphism $\pi_1(X, x) \cong \pi_1(X, y)$.*

**Inner automorphisms.** In particular, a path from $x$ to itself determines an isomorphism of $G = \pi_1(X, x)$ to itself. These are exactly the *inner automorphisms*, of $G$, $g \mapsto hgh^{-1}$.

**Simple connectivity.** We say $X$ is *simply–connected* if it is nonempty, path–connected, and $\pi_1(X, A)$ is trivial. If $X$ is a point, or more generally, if $X$ is contractible, the $\pi_1(X, A)$ is trivial and $X$ is simply–connected. The converse however does not hold – for example the sphere $S^n$ is simply–connected for all $n > 1$, but not contractible.

**Aside: The Poincaré conjecture.** The terminology 'fundamental group' is due to Poincaré. It can be shown that if $X$ is a closed surface, then $X$ is homeomorphic to a sphere $S^2$ iff $\pi_1(X)$ is trivial.

The *Poincaré conjecture* states that if $X$ is a closed 3-manifold, then $X$ is homeomorphic to $S^3$ iff $\pi_1(X)$ is trivial. This statement was finally proved by Perelman in the mid–2000s. It illustrates, perhaps, why this group is so fundamental.

In higher dimensions, one should be careful about whether manifolds and maps are smooth or just topological. Also the fundamental group no longer suffices for a correct formulation. The topological Poincaré conjecture in dimension 4 is true by work of Freedman. The smooth version is still an open question. For dimensions 5 or more, the topological version is true but the smooth version is often false.

**PL theory: taming Peano curves.** A continuous path can be very wild — as we have seen, it can cover an open set in $\mathbb{R}^2$. However if we allow ourselves to change a path by homotopy, then it often can be tamed.

We say a path $f : [a, b] \to \mathbb{R}^n$, is *linear* if it has the form

$$f(t) = a + tb$$

for some $a, b \in \mathbb{R}^n$. A continuous path is *piecewise linear* (PL for short) if we can find $t_0 = a < t_1 < \cdots < t_m = b$ such that $f|[t_i, t_{i+1}]$ is linear for each $i$.

**Theorem 9.3** *Let $f : [0, 1] \to U \subset \mathbb{R}^n$ be a path in an open set in Euclidean space. Then $f$ is homotopic, rel endpoints, to a PL path.*

**Proof.** Since $K = f([0, 1])$ is compact, there is an $r > 0$ such that $B(K, r) \subset U$. By uniform continuity, we can cut $[0, 1]$ into segments $[t_i, t_{i+1}]$ such that the image of each is contained in a ball of radius $\le r$. Now just replace $f$ with the unique function $g$ which is linear on these segments and satisfies $g(t_i) = f(t_i)$. ∎

**Exercise.** The same is true for a path in a finite simplicial complex.

**Functors.** A *functor* is a map that assigns to each object $X$ of a category $\mathcal{C}$ an object $X' = \mathcal{F}(X)$ of a category $\mathcal{C}'$, *and* assigns a morphism $\mathcal{F}(f) = f'$ to every morphism in $\mathcal{C}$. The obvious composition law

$$\mathcal{F}(f \circ g) = \mathcal{F}(f) \circ \mathcal{F}(g)$$

is required to hold; moreover, we require that

$$\mathcal{F}(\mathrm{id}_X) = \mathrm{id}_{\mathcal{F}(X)}\,.$$

The set of path components $\pi_0(X)$ provides a *functor* from the category of topological spaces to the category of sets. In fact it is well–defined for the category whose morphisms are *homotopy equivalent* maps.

The fundamental group is an example of a functor from the category of *pointed* topological spaces $(X, x)$ to the category of groups. In the former category we can take homotopy classes of maps to be the morphisms. In other words, the natural map

$$f_* : \pi_1(X, x) \to \pi_1(Y, y)$$

depends only on the homotopy of $f$ rel basepoints.

In particular this shows:

**Theorem 9.4** *If $X$ and $Y$ are homotopy equivalent and path–connected, then $\pi_1(X)$ is isomorphic to $\pi_1(Y)$.*

# 10 Path lifting and covering spaces

We now come to the first result concerning fundamental groups, which also illustrates many of the general ideals to be developed in the sequel. We will show:

**Theorem 10.1** *The fundamental group of the circle $S^1$ is isomorphic to $\mathbb{Z}$.*

**Covering spaces.** A continuous map $p : E \to B$ *evenly covers* an open set $U \subset B$ if we can write $p^{-1}(U) = \sqcup U_i$ as a union of disjoint open sets, such that $p : U_i \to U$ is a homeomorphism. If every point in $B$ has a neighborhood which is evenly covered by $p : E \to B$, then $p$ is a *covering map* and $E$ is a *covering space* of $B$.

Put differently, once $U$ is small enough, we have a family of continuous branches of the inverse of $p$, $p_i^{-1} : U \to U_i$, and whenever $p(e) \in U$ we have $e \in U_i$ for some $i$.

Here we adopt the standard notation from *fiber bundles* (also used by Munkres), where the *base* of a covering is denoted $B$, and the *covering space* (*éspace étale*) is denote $E$. One could also use $F$ to denote the *fiber*, i.e. the preimage of a point in $B$.

Exercise: if $B$ is connected, then the number of points $d$ in a fiber $F = p^{-1}(p)$ is independent of $p$. The number $d$ is the *degree* of the covering.

**Lifting.** Let $f : X \to B$ be a continuous map. We say $F : X \to B$ is a *lifting* of $f$ if $F$ is also continuous, and $p \circ F = f$. Thus we have a commutative diagram:

$$\begin{tikzcd} & E \arrow[d, "p"] \\ X \arrow[ur, "F"] \arrow[r, "f"] & B. \end{tikzcd}$$

A covering map is *defined* so any lifting problem is *locally solvable*. That is, if $f(X)$ is small enough to be contained in a chart $U$ which is evenly covered, we can simple set $F = p^{-1} \circ U$. The fundamental group enters when we try to do lifting globally.

Note that if we *start* with the map $F : X \to E$, and *define* $f = p \circ F : X \to B$, then $F$ automatically a lift of $f$. In this way we can easily construct lots of solvable lifting problems

To formulate a more precise lifting problem, pick basepoints $x \in X$, $b \in B$ and $e \in E$ such that $f(x) = p(e) = b$, and require that $F(x) = e$. Then a *lifting problem* is to find a continuous map $F$ making the following diagram of pointed spaces commute:

$$\begin{tikzcd} & (E,e) \arrow[d, "p"] \\ (X,x) \arrow[r, "f"] \arrow[ur, dotted, "F"] & (B,b). \end{tikzcd}$$

**Theorem 10.2** *Let $f : [0,1] \to B$ be a path and suppose $b = p(e) = f(0)$. Then $f$ has a unique lifting $F$ such that $F(0) = e$.*

**Proof.** Consider the set $T$ of $t \geq 0$ such that $f|[0,t]$ has a lifting with $F(0) = e$. Because $\pi^{-1}$ is a local homeomorphism, $T$ is open and contains a neighborhood of 0. But $t$ is also closed: if $t_0 = \sup T$, then there is a neighborhood $U$ of $f(t_0)$ that is evenly covered, and if there is an $s \in T$ such that $f([s,t_0]) \subset U$. We can then piece together the lifting of $f|[0,s]$ with a lifting of $f|[s,t_0]$ to get $t_0 \in T$. Thus $f$ has a lifting.

To see the lifting is unique, we observe similarly that if $F'$ is another lifting, then $\{t \,:\, F(t) = F'(t)\}$ is open and closed in $[0,1]$. ∎

**Corollary 10.3** *Suppose $X$ is path connected. Then the solution the lifting is unique, provided it exists.*

**Proof.** Let $F : X \to E$ be a lifting of $f$. Given $x' \in X$, choose a path $\gamma$ from $x$ to $x'$ and lift $f \circ \gamma$ to a path from $e$ to $e'$. This lifting is unique, so it must agree with $F \circ \gamma$. Thus $F(x') = e'$. This shows the value of $F(x')$ can be determined without reference to $F$. ∎

**The circle.** We let $S^1 = \mathbb{R}/\mathbb{Z}$. Thus a point $x \in S^1$ is a real number modulo 1. There is a natural action of $\mathbb{R}$ on $S^1$ by rotation, given by $x \mapsto x + \theta \bmod 1$.

We define a metric on $S^1$ by

$$d(x,y) = \inf\{|\theta| \,:\, \theta \in \mathbb{R},\ x + \theta = y\}.$$

This is the minimum distance between $x$ and $y$ in the natural metric that gives the circle unit length. Thus $d(x,y) \leq 1/2$ for all $x, y \in S^1$.

We now have a natural covering map

$$p : \mathbb{R} \to S^1,$$

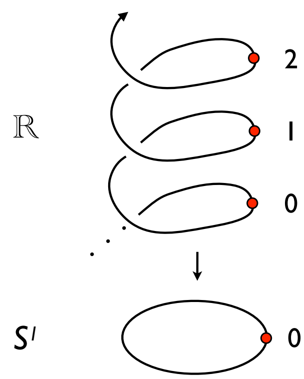



Figure 5. The covering map $\mathbb{R} \to S^1$ can be pictured as a parking garage.

give by $p(x) = x \bmod 1$. See Figure 5.

In this setting, it is easy to understand path lifting. Namely if $f([0,1])$ lies entirely in an interval $I \subset S^1$, then $p^{-1}$ has a single valued branch along $I$ and we can simply set $F = p^{-1} \circ f$. The branch is uniquely determine once we specify $F(0)$. By uniform continuity, any path $f : [0,1] \to S^1$ can be broken up into pieces of this type; by lifting each one, we get a lift of $f$.

**Space of paths and loops.** The space of paths $\mathcal{P}_b(B)$ consists of all $f : ([0,1], 0) \to (B, b)$. If $B$ is a metric space then so is $\mathcal{P}_b(B)$, with the metric
$$D(f,g) = \sup_t d(f(t), g(t)).$$
Indeed, $\mathcal{P}_b(B)$ is a closed subspace of $C([0,1], B)$.

Similarly, if we impose the condition that $f(1) = b$ as well, we obtain the space of loops $\Omega_b(B)$.

**Continuity of lifting.** Now consider the case of the circle. Path lifting defines a natural map
$$L : \mathcal{P}_0(S^1) \to \mathcal{P}_0(\mathbb{R}).$$

**Theorem 10.4** *Path lifting is continuous.*

**Proof.** We claim that if $D(f,g) < 1/100$ then $D(F,G) = d(f,g)$. Indeed, for each $t \in [0,1]$, either $d(F(t),G(t)) = d(f(t),g(t))$, or $d(F(t),G(t)) \geq 1/2$. Both sets are closed, and the first contains $t = 0$, so it contains all $t \in [0,1]$. Thus path lifting is actually a local isometric from $\mathcal{P}_0(S^1)$ to $\mathcal{P}_0(\mathbb{R})$. ■

**Corollary 10.5** *If $f_0, f_1 \in \mathcal{P}_0(S^1)$ are homotopic, then so are their lifts.*

**Proof.** The maps $F_t = L(f_t)$ give a homotopy between their lifts. ∎

Note: By the definition of $L$, the value $F_t(0) = 0$ is constant for the lifted homotopy. If $f_t(1)$ is constant during the homotopy from $f_0$ to $f_1$, then the same is true for $F_t(1)$.

**The fundamental group.** Now let $\Omega_0(S^1)$ be the space of *loops*, i.e. paths such that $f(0) = f(1) = 0$. If $f \in \Omega_0(S^1)$ and $F = L(f)$, then $F(1) \in p^{-1}(0) = \mathbb{Z}$. Thus we have a *continuous* map

$$T : \Omega_0(S^1) \to \mathbb{Z}.$$

Since $T(f)$ is continuous, it only depends on the *homotopy class* of $f$. Thus $T$ descends to a map

$$T : \pi_1(S^1, 0) \to \mathbb{Z}.$$

**Theorem 10.6** *The map $T$ is an isomorphism.*

**Proof.** Suppose $T(f) = 0$. Then $F = L(f)$ lies in $\Omega_0(\mathbb{R})$. In other words, $[f] = p_*[F]$. Since $\pi_1(\mathbb{R}, 0)$ is trivial, $[f]$ is the identity element in $\pi_1(S^1, 0)$. Thus $T$ is injective.

Taking $F_n(t) = nt$ and $f = p \circ F$, we see that $T$ is surjective. ∎

**Degree.** In general we let $[X, Y]$ denote the set of homotopy classes of maps $f : X \to Y$. For the circle, any $f \in [S^1, S^1]$ determines a map $f_* : \pi_1(S^1) \to \pi_1(S^1)$ which is completely determined by the integer

$$\deg(f) = (f_*)(1) \in \mathbb{Z}.$$

Conversely, two maps of the same degree are homotopic. Apart from shifting around basepoints, this statement is equivalent to the assertion that $\pi_1(S^1) \cong \mathbb{Z}$.

Put differently, we have a continuous map

$$\deg : [S^1, S^1] \to \mathbb{Z},$$

and the fibers of this map are path–connected. That is, $[S^1, S^1]$ breaks up into components that are classified by the degree of the map.

For a third perspective, suppose $f : S^1 \to S^1$. By path lifting, we can then find a map $F : \mathbb{R} \to \mathbb{R}$ such that $p(F(x)) = f(p(x))$. This map satisfies

$$F(x+1) = F(x) + \deg(f).$$

Explicitly, we can consider $f$ has a map from $[0,1]$ to $S^1$, and then lift it to a map $F : [0,1] \to [x, x+\deg(f)]$ for some $x$. We then extend $F$ periodically to $\mathbb{R}$.

**Degree and preimages.** Since $S^1$ is modeled on $\mathbb{R}$, it makes sense to talk about smooth or piecewise–linear maps $f : S^1 \to S^1$. For example $f'(x)$ makes sense for any smooth map $f : S^1 \to S^1$, since the derivative $F'(x)$ of any lift of $f$ is invariant under $x \mapsto x+n$.

Now let $\deg(f,x) = \operatorname{sign} f'(x) = -1, 0$ or 1. We set $\deg(f,x) = 0$ if $f'(x)$ does not exist. The *singular points* of $f$ are the $x$ with $\deg(f,x) = 0$; their images are the singular values.

**Theorem 10.7** *Let $f : S^1 \to S^1$ be a smooth or PL map and suppose $y \in S^1$ is not a singular value of $f$. Then $f^{-1}(y)$ is finite, and* $\deg(f) = \sum_{f(x)=y} \deg(f,x)$.

**General path lifting.** The same results hold for general covering spaces. First, as we have seen, any path $f : [0,1] \to B$ with $b_0 = f(0) = p(e_0)$ has a unique lift $F$ with $F(0) = e_0$. More generally, given

$$f : [0,1]^n \to B$$

with $f(0,\ldots,0) = b_0 = p(e_0)$, there is a unique lift

$$F : [0,1]^n \to E$$

satisfying $F(0,\ldots,0) = e_0$. In other words we have:

**Theorem 10.8** *Any map $f : [0,1]^n \to (B,b)$ has a unique lift $F : [0,1]^n \to (E,e)$.*

**Proof.** The proof is clear if $f(I^n)$ $B$ itself is evenly covered. For in this case, $E = \sqcup B_i$ and we have inverse homeomorphisms $p_i : B \to B_i$ for each $i$. There is a unique $i$ such that $e \in B_i$, and the unique lifting is given by $F = p_i \circ f$. To see it is unique, observe that any other lift still satisfies $F(0) = e \in B_i$, so

$F$ sends $I^n$ into $B_i$ because $I^n$ is connected. Since $p \circ F = f$, we must have $F = p_i \circ f$.

For the general case, cut $I^n$ into cubes that are small enough that they are evenly covered, and then lift them one at a time. The key point is that cubes can be ordered so one never runs into consistent difficulties with lifts. This reflects the fact that $I^n$ is simply–connected. ∎

**Corollary 10.9** *Homotopic paths have homotopic lifts.*

**Proof.** Lift the homotopy. ∎

# 11 Global topology: applications

Here are some remarkable applications of the fact that $\pi_1(S^1) \cong \mathbb{Z}$. They are all, so to speak, advanced versions of the intermediate value theorem, which itself comes from the fact that $\pi_0(I) = 0$.

Let $B^n$ denote the *closed* unit ball in $\mathbb{R}^n$, and $S^{n-1} \subset B^n$ the unit sphere. The *antipodal map* of the sphere is given by $x \mapsto -x$ in these coordinates. Note that the antipodal map has no fixed points (it should not be confused with the map $x \mapsto -x \bmod 1$ on $S^1$.)

1. There is no retraction of $B^2$ onto $S^1$.
2. (Brouwer) Any continuous map $F : B^2 \to B^2$ has a fixed point.
3. (Perron–Frobenius) Any 3x3 matrix with $A_{ij} > 0$ has a positive eigenvector.
4. (Gauss) Every nonconstant complex polynomial has a root in $\mathbb{C}$.
5. If $f : S^1 \to S^1$ satisfies $f(-x) = -f(x)$, then $\deg(f)$ is odd.
6. If $F : S^2 \to \mathbb{R}^2$ is continuous, then $F(x) = F(-x)$ for some $x \in S^2$.
7. (Borsuk–Ulam) Given 2 pieces of cheese on a cutting board, there is a single stroke of the knife that will cut both in half.

We now turn to the proofs.

**1. Retraction.** Suppose we have a retraction $f : B^2 \to S^1$. Let $i : S^1 \to S^1$ be the inclusion. Then $g = f \circ i$ is the identity, so $g_*$ gives the identity on $\pi_1(S^1) \cong \mathbb{Z}$. But this map can be factored as $g_* = f_* \circ i_*$ where

$$\mathbb{Z} \cong \pi_1(S^1) \xrightarrow{i_*} \pi_1(B^2) = 0 \xrightarrow{f_*} \pi_1(S^1) \cong \mathbb{Z},$$

so $g_* = 0$. This is a contradiction.

Alternatively we note:

> A continuous map $f : S^1 \to S^1$ extends to a map $B^2 \to S^1$ if and only if $f$ has degree zero.

**2. Fixed points.** The Brouwer theorem is true for $B^n$, not just $B^2$. Note that the case of $B^1$ does in fact follow from the intermediate value theorem.

Now suppose $F : B^2 \to B^2$ has no fixed points. We then get a map $\rho : B^2 \to S^1$ by following the ray from $F(x)$ to $x$ until it hits $\partial B^2$. Then $F|S^1$ is the identity. Hence $B^2$ retracts onto $S^1$, which is impossible. Thus $F$ has a fixed point!

Brouwer famously argued, in founding *intuitionism*, that the law of the excluded middle (proof by contradiction) is mathematically unsound, thereby invalidating his own proof.

**3. Eigenvectors.** The map $A$ sends the closed simplex $S$ of vectors $v_i$ with $\sum v_i = 1$ into itself, by $F(v) = A(v)/(\sum(Av)_i)$. Since $S$ is homeomorphic to $B^2$, $F$ has fixed point, and hence an eigenvector. Noting that $A(v)$ is *strictly positive* for all $v \in S$, we see the eigenvector is positive.

Note: this case is actually not so hard: the map $F : S \to S$ turns out to be a contraction for a suitable metric, so the fixed point can be found by iteration.

**4. Zeros.** Suppose $F(z) = z^d + a_1 z^{d-1} + \cdots + a_d$ is a polynomial with no zero. Then for every radius $r \geq 0$ we have a map $f_r : S^1 \to S^1$ given by

$$f_r(z) = F(rz)/|F(rz)|.$$

These maps vary continuously with $r$. Since $f_0$ is constant, $\deg(f_r) = 0$ for all $r$. But $f_r(z) \to z^d$ uniformly as $r \to \infty$. Thus $\deg(f_r) = d > 0$ for all $r$ sufficiently large, which is a contradiction.

**5. Antipodes.** Suppose $f : S^1 \to S^1$ satisfies $f(-x) = -f(x)$. We can then lift $f$ to a map $F : \mathbb{R} \to \mathbb{R}$ such that $F(x + 1) = F(x) + \deg(f)$, and

$F(x + 1/2) = F(x) + e(x)$, where $e(x) \in 1/2 + \mathbb{Z}$. Since $e(x)$ is continuous, it must be constant, and since $F(x+1) = F(F(x+1/2)+1/2) = F(x)+2e$, we must have $e = \deg(f)/2$. Thus $\deg(f)$ is odd.

**6. Squashed spheres.** Suppose $F : S^2 \to \mathbb{R}^2$ satisfies $F(x) \neq F(-x)$ for all $x \in S^2$. We can then define a map $\rho : S^2 \to S^1$ by

$$\rho(x) = \frac{F(x) - F(-x)}{|F(x) - F(-x)|}.$$

Naturally $\rho|S^1$ has degree zero, since it extends over the northern hemisphere $B^2$. But $\rho(-x) = -\rho(x)$, so $\rho|S^1$ must have odd degree, a contradiction.

This result also holds for $S^n$.

**7. Sandwich preparation.** The space of oriented lines in the plane — with a pair of lines at infinity added — is homeomorphic to $S^2$. This can be seen in several ways. The intrinsic geometric approach is to observe that $\mathbb{RP}^2 = \mathbb{R}^2 \cup \mathbb{RP}^1$ is homeomorphic to $S^2/(x \mapsto -x)$, and that lines and points are in projective duality. More naively, think of the plane as $P = \mathbb{R}^2 \times \{2\} \subset \mathbb{R}^3$, and think of $S^2$ as the unit sphere centered at 0. Then any $x \in S^2$ defines a plane $x^\perp$ which meets $\mathbb{R}^2$ in a line with a natural orientation, *unless* $x$ is the north or south pole of $S^2$. Conversely, any line $L \subset P$ can be extended to a plane $P$ through the origin in $\mathbb{R}^2$, and an orientation of $L$ determines one of the points in $P^\perp \cap S^2$.

Now let $A_1, A_2$ be two bounded regions of positive area in $P$. Define a map $F : S^2 \to \mathbb{R}^2$ by taking the halfspace $H_x \subset P$ defined by $x \in S^2$, and setting

$$F(x) = (\mathrm{area}(A_1 \cap H_x), \mathrm{area}(A_2 \cap H_x)).$$

This makes sense even at the north and south poles of $S^2$. By the previous result, there is an $x$ such that $F(x) = F(-x)$. But this means exactly that $L$ cuts both $A_1$ and $A_2$ in half.

**The ham sandwich theorem.** Here is an extension to higher dimensions.

**Theorem 11.1** *Given 3 bounded open sets $A$, $B$ and $C$ in $\mathbb{R}^3$, there exists a plane that cuts each of them in half.*

**Proof.** After scaling we can assume $A$, $B$ and $C$ are all subsets of the unit ball $B^3$. Now for each $x \in S^2$ we consider the line $L_x$ from $x$ to $-x$. Along this line we have three naturally distinguished points: $x_A$, $x_B$, $x_C$ — such that the plane perpendicular to $L_x$ through $x_A$ cuts $A$ in half, and similarly for $x_B$ and $x_C$. Define a map $f : S^2 \to \mathbb{R}^2$ by

$$f(x) = (\langle x, x_A - x_B\rangle, \langle x, x_A - x_C\rangle).$$

Note that $f(-x) = -f(x)$. We must have a point such that $f(x) = f(-x)$; at this point $x_A = x_B = x_C$, and so the plane normal to $L_x$ at this point bisects $A$, $B$ and $C$ simultaneously. ∎

# 12 Quotients, gluing and simplicial complexes

In this section we discuss the idea of gluing together spaces out of simpler pieces, and especially simplicial complexes, to give some interesting low–dimensional examples.

**Gluing and quotients.** Let $X$ be a topological space, let $A$ be a set and let $f : X \to A$ be a surjective map. The *quotient topology* on $A$ is defined by

$$U \subset A \text{ is open } \iff f^{-1}U \subset X \text{ is open.}$$

(Exercise: check this is really a topology).

A map $f : X \to Y$ is called a *quotient map* if $f$ is surjective and $f^{-1}(U)$ is open iff $U$ is open in $Y$. With the quotient topology on $A$, $f : X \to A$ becomes a quotient map. Moreover, given any surjective map $f : X \to Y$, there is a *unique* topology on $Y$ making $f$ a quotient map.

The set $A$ can always be constructed by introducing an equivalence relation $\sim$ on $X$, and setting $A = X/\sim$. (The relation is $x \sim x'$ iff $f(x) = f(x')$.) Often this equivalence relation is described by a homeomorphism $f : B \to C$ between disjoint of $X$. Then we declare $x \sim y$ if $f(x) = y$, and refer to $A = X/\sim$ as the result of 'gluing $A$ to $B$ with $f$.'

**Simplicial complexes.** Here is another general language for describing spaces combinatorially.

Let $V$ be a finite set, and let $\mathcal{S} \subset \mathcal{P}(V)$ satisfying the saturation property:

$$A \subset B \in \mathcal{S} \implies A \in \mathcal{S}.$$

Let $\Delta_A \subset \mathbb{R}^V$ be defined as the set of vectors with $\sum v_i = 1$ and $v_i = 0$ unless $i \in A$. Clearly $\Delta_A$ is homeomorphic to a simplex, and $\Delta_A \subset \Delta_B$ if $A \subset B$.

The data $X = (V, \mathcal{S})$ give a *simplicial complex*, realized by the topological space

$$|X| = \bigcup_{A \in \mathcal{S}} \Delta_A.$$

**Graphs.** A graph $G$ with vertices $V(G)$ and edges $E(G) \subset \mathcal{P}_2(V)$, and no loops or parallel edges, is a simplicial complex of dimension 1; conversely, any 1-dimensional simplicial complex can be considered a graph. The cell structure is part of the data, in both cases.

We can also regard the topological space underlying a graph as a quotient space:

$$X = V(G) \sqcup E(G) \times [0, 1]$$

with the obvious identifications.

**The circle.** We can think of $S^1$ as $[0, 1]$ with 0 glued to 1. Alternatively it can be realized as the $1 - \textit{skeleton}$ of a 2-simplex. (The same thing works for the $n$-sphere.)

**Wedges of circles.** The space $(X, x) = \vee_1^n (X_i, x_i)$ is obtained from $\sqcup X_i$ by identifying $A = \{x_1, \ldots, x_\}$ to a single point $x$. If all the factors $X_i$ are homogeneous, we often omit the $x_i$.

**Theorem 12.1** *Any finite graph is homotopy equivalent to a bouquet of circles,* $\vee_1^n S^1$.

**Proof.** Let $T \subset G$ be a maximal tree. Then $T$ is contractible, and $G/T$ is a wedge of circles. ∎

**The bridges of Königsberg.** One of the first 'purely topological' problems studied by Euler (1736) has to do with the weekly *Sonntagbummel* in the Prussian city of Königsberg (now the a Russian enclave). Residents tried to cross all the bridges over the river Pregel, but no bridge twice. They failed! Why? (Note: Kant was born in Königsberg (1724).)

**The letter $\boldsymbol{Y}$.** Here is a problem at the boundary of combinatorial and point set topology: show you cannot embed uncountably many disjoint copies of the letter $Y$ in the plane.

**Examples from the square.** There are several interesting spaces $Y$ that can be constructed from the square $X = [0, 1]^2$ with various identifications. To describe these, let $A = [0, 1] \times \{0\}$ and $A' = \{1\} \times [0, 1]$ be the bottom

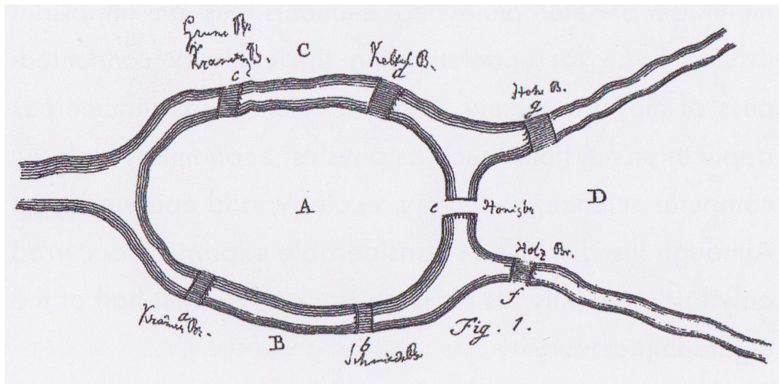



Figure 6. The bridges of Königsberg.

and top of the square, and let $B$ and $B'$ be the right and left sides. We orient all these so the $x$ and $y$ coordinates are increasing on them. An edge $J$ with reversed orientation will be written $-J$.

1. *The cylinder.* If we glue $B$ to $B'$ we obtain a cylinder, isomorphism to $S^1 \times [0,1]$. A typical neighborhood of a point on the gluing line corresponds to two half-moons along $B$ and $B'$ in $X$. The same space results if we glue $A$ to $A'$.

   Note that the horizontal lines in $X$ become parallel closed circles in $Y$.

2. *The Möbius band.* If we glue $B$ to $-B'$ the result is a twisted cylinder called the Möbius band. The horizontal lines again become circles, but they behave very differentially now! Most of the circles have length 2, but the core of the Möbius band has length 1. The outer circle projects to the central circle by a degree two covering map.

3. *The cone.* If we glue $A$ to $B$, we get a topological cone on $S^1$.

4. *The mystery space.* What space do we get if we glue $A$ to $B'$? This example is tricky. Note that the three vertices in $A \cup B'$ are identified to a single point.

   Hint: insert a new edge $C$ between $A$ and $B'$, to obtain a pentagon with two disjoint edges glued together. What space $X$ results from gluing up $A$ and $B'$ now? What does the image $C'$ of $C$ on $X$ look like? What is $X/C'$?

5. *The torus.* If we glue $A$ to $A'$ and $B$ to $B'$, the result is homeomorphic to $S^1 \times S^1$. The horizontal lines in $X$ give a family of parallel circles on $Y$.

6. *The sphere.* If we glue $A$ to $B$ and $A'$ to $B'$, we get a space homeomorphic to $S^2$. It is useful to use coordinates on $S^2$ given by latitude $\alpha \in [-\pi/2, \pi/2]$ and longitude $\theta \in S^1 = \mathbb{R}/2\pi\mathbb{Z}$. The northern hemisphere $N$ is given by $\alpha \geq 0$, and the western hemisphere $W$ by $0 \leq \theta \leq \pi$.

   To see that $Y \cong S^2$ in detail, first we project $X$ to $N$. Then $Y$ is homeomorphic to $N$ with $(0, \theta)$ glued to $(0, -\theta)$ for all $\theta$. Now we rotate to $N$ to $W$. Then the relation becomes $(0, \theta) \sim (\pi, \theta)$. Finally we observe that the quotient map $f : W \to S^2$ given by $f(\alpha, \theta) = (\alpha, 2\theta)$ makes exactly these identifications.

7. *The Klein bottle.* Now glue $B$ to $B'$ but glue $A$ to $-A'$. The result is a closed, nonorientable surface! Unlike the torus, it cannot be embedded in 3-space.

   The way a physical Klein bottle is usually constructed, it passes through itself in a way that abstract Klein bottle does not. One can by analogy consider how a Möbius band might be represented in flatland: it is a strip in the plane that passes through itself as well.

8. *The real projective plane* $\mathbb{RP}^2$. Probably the least familiar space is obtained by gluing $A$ to $-A'$ and $B$ to $-B'$. This is equivalent to taking the quotient of the closed unit disk $D \subset \mathbb{R}^2$ by the relation $v \sim -v$ on the circle.

9. *The dunce cap.* There are also a few interesting spaces that can be built from a triangle, the most famous of which is the dunce cap. Here we identify $A$ to $B$ and $-C$. Remarkably, the resulting space $X$ is contractible (but not collapsible!).

**Some assembly required.** All of these spaces are simplicial complexes, however some subdivision may be required to see this. For example, the dunce hat is not a simplicial complex with the single 2-cell given, but after subdivision it becomes a simplicial complex.

**Quotients by group actions.** Group actions provide another important source of quotient spaces.

Let $G$ be a group acting on a topological space $X$. Two points lie in the same *orbit* if $G \cdot x$ $G \cdot y$. We let $Y = X/G$ denote the space of orbits. There is a natural map $X \to X/G$, and we give $X/G$ the quotient topology.

We say $G$ acts *evenly* if any $x \in X$ has a neighborhood $U$ such that $U \cap g(U) = \emptyset$ whenever $g \neq \mathrm{id}$. In this case $X \to X/G$ is a covering map. In general we allow nontrivial elements of $G$ to have fixed points, and then the covering property is violated.

**Examples.** We have $S^1/\langle g \rangle = [0,1]$ where $g(z) = \overline{z}$. We have $S^2/\langle g \rangle = S^2$ where $g$ is a finite–ordered rotation. We have $S^2/\langle g \rangle = B^2$ where $g$ is reflection through the equator. None of these are covering spaces.

We have $\mathbb{R}/\mathbb{Z} \cong S^1$. We have $\mathbb{RP}^2 = S^2/\langle \alpha \rangle$. The torus is $\mathbb{R}^2/\mathbb{Z}^2$, or more generally $\mathbb{R}^2/\Lambda$ for any lattice $\Lambda$. All of these are covering spaces.

**The projective plane and the Möbius band.** The complement of a ball in $X = \mathbb{RP}^2$ is homeomorphic to the Möbius band. To see this, think of $X$ as $S^2/\langle \alpha \rangle$. Then $X$ minus a disk is the quotient of a band around the equator by the antipodal map.

**Distinguishing by fundamental group.** We can compute $\pi_1(X)$ for all of these spaces without too much effort. E.g. we get $\mathbb{Z}$ for the cylinder and the Möbius band (in fact these spaces are both homotopy equivalent to $S^1$), and $\mathbb{Z} \times \mathbb{Z}$ for the torus. The real projective plane has $\pi_1(X) \cong \mathbb{Z}/2$; this is our first example of torsion in the fundamental group. The Klein bottle is the most interesting case.

**Fundamental group of the dunce cap.** This space illustrates some important principles. First, we define the $k$-skeleton $X^k$ of a simplicial complex to be the union of its $k$-cells.

**Theorem 12.2** *If $X$ is a connected cell complex, then the map $\pi_1(X^1) \to \pi_1(X)$ is surjective.*

**Proof.** We may choose our basepoint $x$ in $X^1$. Any loop $f : [0,1] \to X$ based at $x$ can be homotoped, rel endpoints, to a PL path which misses the barycenter $b_\Delta$ of every simplex $\Delta$ of dimension 2 or more. Then projection from $b_\Delta$ gives a homotopy from $f$ to a path in $X^1$. ∎

**Corollary 12.3** *The dunce cap is simply–connected.*

**Proof.** In this case $X^1$ is a circle. There is a natural 'attaching map' $f : S^1 \to X^1$ which winds positively twice and then negatively once. This map extends to the 2-ball by the definition of $X$. Thus $\pi_1(X)$ is generated by a loop which is homotopically trivial. ■

# 13 Galois theory of covering spaces

In this section we show that the lifting problem for general covering spaces can be solved iff the corresponding problem in group theory has a solution. We then use this fact to classify the tower of covering spaces over a given base, in a manner that parallels the classification of field extension in Galois theory.

**Lifting general maps.** Let $\pi : E \to B$ be a covering map. To discuss the lifting problem more precisely, we introduce basepoints everywhere. The notation $f : (A, a) \to (B, b)$ means $f : A \to B$ is a continuous map and $b = f(a)$. The category of pointed spaces has already occurred as the natural domain of the *functor* $\pi_1$.

A lifting problem is described by the following diagram.

$$\begin{tikzcd} & (E,e) \arrow[d, "p"] \\ (X,x) \arrow[ur, dotted, "F"] \arrow[r, "f"] & (B,b). \end{tikzcd}$$

We are given $p$ and $f$, and we wish to lift $f$ to a map $F$ with $F(x) = e$ a given point above $b$.

In this setting, it is not difficult to show that a lifting is *unique* if it exists. But what about existence?

A *necessary* condition is obtained by applying the fundamental group functor. That is, we have an associated problem in *group theory*:

A lifting problem is described by the following diagram.

$$\begin{tikzcd} & \pi_1(E,e) \arrow[d, "\pi_*"] \\ \pi_1(X,x) \arrow[ur, dotted, "F_*"] \arrow[r, "f_*"] & \pi_1(B,b). \end{tikzcd}$$

Clearly it is necessary that the group theory problem have a solution for the original lifting problem to have a solution. In this section we will show the converse.

A special case, which we have already demonstrated, arises when $X = I^n$. Then $\pi_1(X, x)$ is trivial and the lifting problem can always be solved, as we have seen. A consequence of this fact is:

**Lemma 13.1** *The map $p_* : \pi_1(E, e) \to \pi_1(B, b)$ is injective.*

**Proof.** Let $F_0 : [0, 1] \to E$ be a loop based at $e$, and suppose $f_0 = p \circ F_0$ is trivial in $\pi_1(B, b)$. Let $f_t$ be a deformation of $f_0$ to the constant path $f_1$. Then $f_t$ lifts to a deformation $F_t$ of $F_0$ to the constant path $F_1$. (One may check that $F_t$ is a loop for all $t$, since $F_0$ is.) This shows that the kernel of $p_*$ is trivial, and hence $p_*$ is injective. ∎

Because of this Lemma, we may regard $\pi_1(E, e)$ as a *subgroup* of $\pi_1(B, b)$.

**Lemma 13.2** *If $f : [0, 1] \to B$ is a loop based at $b$ and $[f] \in \pi_1(E, e)$, then $f$ lifts to a loop in $E$ based at $e$.*

**Proof.** From the definitions, $f$ is homotopic to a loop $g$ that lifts. But the homotopy also lifts, so $f$ lifts. ∎

**Theorem 13.3** *Suppose $X$ is path–connected and locally path–connected. Then the lifting problem has a solution iff $f_*$ sends $\pi_1(X, x)$ into $\pi_1(E, e)$.*

**Proof.** If the lifting problem can be solved, then $f_* = p_* \circ F_*$ and so $f_*$ sends $\pi_1(X, x)$ into $\pi_1(E, e) = \mathrm{Im}(p)$.

For the converse, consider any point $y \in X$. Since $X$ is path–connected, we may choose a path $g : [0, 1] \to X$ connecting $x$ to $y$. Lift $f \circ g$ to a path $G$ in $E$ starting at $e$. We define $F(y)$ to be the endpoint of this lifted path. Then certainly $p \circ F(y) = f(y)$.

We must check that $F(y)$ is well-defined. But if we have another path $h$ from $x$ to $y$, then we can put the two together to get a loop in $(X, x)$. The image of this loop lifts to $(E, e)$ by the assumption that $\pi_1(X, x)$ maps into $\pi_1(E, e)$. This means the two lifts agree, so they piece together upstairs in $E$.

Finally we must check that $F$ is continuous. Here we use the fact that $y$ has a path–connected neighborhood $U$. Shrinking $U$ if necessary, we can assume that $f(U)$ is evenly covered by $p$. This means we have a continuous branch $p^{-1}$ of $p$ on $f(U)$. Now for any $y' \in U$ we can choose $g' = h * g$, where $h$ is a path from $y$ to $y'$ in $U$. Then $\pi^{-1} \circ f \circ h$ provides a lift of $f \circ h$ to a path from $F(y)$ to $F(y')$. But clearly $F(y') = \pi^{-1}(f(y'))$, so $F$ is continuous.

■

**Examples from complex analysis.**

1. The map $\exp : \mathbb{C} \to \mathbb{C}^*$ is a covering map, so the results above show:

   *There is a continuous, analytic branch of* $\log(z)$ *on any simply–connected region* $U \subset \mathbb{C}^*$*, or more generally on any region that does not contain a loop around the origin.*

   This branch is uniquely determined once $\log(p)$ is specified for one $p \in U$.

   In fact, $\log(z) = \log|z| + i \arg(z)$ is only multivalued because the argument of $z$ is only well-defined modulo $2\pi$. Thus the key fact is that the argument of $z$ can be defined on such a region.

2. Let $f(z)$ be a polynomial, and let $U \subset \mathbb{C}$ be a simply–connected region disjoint from the zeros of $p$. Then a single–valued branch of $f(z)^{1/d}$ can be defined on this region.

   In fact, if we let $Z(f)$ denote the zeros of $f$, then we have a map $f : \mathbb{C} - Z(f) \to \mathbb{C}^*$ and a covering map $p : \mathbb{C}^* \to \mathbb{C}^*$ given by $p(z) = z^d$. Then the lifting problem can be solved since $\pi_1(U)$ is trivial.

3. More generally, we can ask, given an arbitrary connected region $U \subset \mathbb{C} - Z(f)$, when can $f(z)^{1/d}$ be defined consistently throughout $U$? For this, it is necessary and sufficient that

   $$f_*(\pi_1(U)) \subset d\mathbb{Z} \subset \mathbb{Z} \cong \pi_1(\mathbb{C}^*).$$

   For example, if $f$ has simple zeros, and each bounded component of $\mathbb{C} - U$ contains an even number of zeros, then $f(z)^{1/2}$ can be consistently defined. Thus if $f(z) = (z - a_1) \cdots (z - a_{2n})$ with $a_i \in \mathbb{R}$, we can define $\sqrt{f(z)}$ on $U = \mathbb{C} - J$ where $J = [a_1, a_2] \cup \cdots \cup [a_{2n-1}, a_{2n}]$.

**Second proof of oddness of degree.** Here is a proof using lifting that a map $f : S^1 \to S^1$ with $f(-z) = -f(z)$ has odd degree. Think of $S^1$ as the circle in the complex plane, and let $p(z) = z^2$. Then $p : S^1 \to S^1$ is a covering map, with $\operatorname{Im} p_* = 2\mathbb{Z} \subset \mathbb{Z}$, and $p(r(z)) = p(z)$.

Suppose $f(-z) = -f(z)$. Then $F(z) = f(\sqrt{z})^2$ is well–defined, which means we have a function $F : S^1 \to S^1$ such that the diagram:

$$\begin{tikzcd} S^1 \arrow[r, "f"] \arrow[d, "p"] & S^1 \arrow[d, "p"] \\ S^1 \arrow[r, "F"] \arrow[ur, dotted, "g"] & S^1 \end{tikzcd}$$

commutes. Clearly $\deg(f) = \deg(F)$. Now if $\deg(F)$ is even, then $F$ has a lift $g$ as above. We now have to be careful, because we only know at first that the lower triangle is commutative: $p \circ g = F$. But this gives:

$$g(z^2)^2 = F(z^2) = f(z)^2.$$

Thus $g(z^2)/f(z) = \pm 1$ for each $z$. Since this function is continuous, it has a constant value $\epsilon = \pm 1$. Thus $f(z) = \epsilon g(z^2)$. But then $f(z) = f(-z)$, contrary to out assumption.

**Classification of covering spaces.** Let $B$ be a reasonable connected space, for example a simplicial complex or an open subset of $\mathbb{R}^n$. By reasonable we mean that $B$ is path–connected, locally path–connected, and locally simply–connected. We wish to classify covering spaces $E$ of $B$ up to isomorphism over $B$.

We also wish to study the *Galois group* of a covering. For $p : E \to B$ the group $\operatorname{Gal}(E/B)$ consists of the homeomorphisms $f : E \to E$ that satisfy $p \circ f = p$. Note that this group does *not* depend on a choice of basepoint. We say $E/B$ is a *normal* or regular or Galois covering if $\operatorname{Gal}(E/B)$ acts transitively on the fibers of $E/B$.

We say covering spaces $E$ and $E'$ are isomorphic *over* $B$ if there exists a homeomorphism $h : (E, e) \to (E', e')$ such that $p = p' \circ h$. If we only require that $h$ is continuous, then we say $E$ *factors through* $E'$.

Here is a quick summary of the classification.

1. For any covering space $p : (E, e) \to (B, b)$ we get an injection

   $$p_* : \pi_1(E, e) \to \pi_1(B, b),$$

   so $E$ determines a subgroup $H_E \subset G = \pi_1(B, b)$. (See Lemma 13.1).

2. Under the natural map $E \mapsto H_E = \pi_1(E,e) \subset G = \pi_1(B,b)$:

   *Covering spaces of $(B,b)$, up to isomorphism over $(B,b)$, correspond bijectively to subgroups of $G = \pi_1(B,b)$.*

3. A covering $E$ factors through $E'$ iff $H_E \subset H_{E'}$. (These last two points follow from Theorem 13.3.)

4. The covering space $E$ is normal iff $H_E$ is a normal subgroup of $G$, in which case
$$\mathrm{Gal}(E/B) \cong G/H_E.$$
   (This isomorphism is easy to construct: given a deck transformation $g$, connected $e$ to $g(e)$ and take its image in $\pi_1(B,b)/\pi_1(E,e)$.)

5. The space $B$ has a canonical *universal cover* $p : \widetilde{B} \to B$ such that $\pi_1(\widetilde{B})$ is trivial and
$$\mathrm{Gal}(\widetilde{B}/B) \cong \pi_1(B,b).$$
   (See Theorem 13.7 below).

6. The covering space $E = H\backslash\widetilde{B}$ has
$$H = \pi_1(E,e) = \mathrm{Gal}(\widetilde{B}/E) = H_E \subset G = \pi_1(B,b).$$

**Classification of covering spaces: details.** Using lifting we quickly see the main point:

**Theorem 13.4** *The covering $E$ factors through $E'$ iff $H_E \subset H_{E'}$.*

**Proof.** To see that covering factors is exactly to say that $p : (E,e) \to (B,b)$ lifts to $(E',e')$, and this happens iff $H_E \subset H_{E'}$ by our previous lifting theorem. ∎

If there is a factorization both ways, then $E \cong E'$ over $B$.

**Theorem 13.5** *If $E$ is the universal cover of $B$, then* $\mathrm{Gal}(E/B) \cong \pi_1(B)$.

**Proof.** Choose basepoints with $p(e) = b$, and let $F = \pi^{-1}(b)$. Then $(E, e) \cong (E, e')$ for any $e' \in F$. This isomorphism gives an element of the deck group. Conversely, an element $g \in \mathrm{Gal}(E/B)$ is uniquely determined once we know $e' = g(e)$. Thus $\mathrm{Gal}(E, B)$ is naturally identified with the set $F$.

Now the set $F$ is naturally identified with $\pi_1(B, b)$, since a path from $e$ to $e' \in F$ projects to a loop in $B$, and any two such paths are homotopic since $E$ is simply–connected. Thus we have a natural bijection $\phi : \mathrm{Gal}(E/B) \to \pi_1(B, b)$. It remains only to prove that $\phi$ is a homomorphism.

To this end, consider a composition $[f_1 * f_2] \in \pi_1(B, b)$, with lifts $\widetilde{f_i}$ connecting $e$ to $g_i(e)$, $i = 1, 2$. Then the lift of $f = f_1 * f_2$ based at $e$ is given by

$$\widetilde{f} = \widetilde{f_1} * (g_1 \cdot \widetilde{f_2}),$$

which satisfies $\widetilde{f}(1) = g_1(\widetilde{f_2}(1)) = g_1(g_2(e))$. Thus $\phi$ is an isomorphism of groups. ∎

**Example: $S^1$.** The simplest example arises when $B = S^1$. Then $\pi_1(B, b) \cong \mathbb{Z}$. The trivial subgroup corresponds to the contractible cover, $p : \mathbb{R} \to S^1$. The subgroup $d\mathbb{Z}$ corresponds to the covering map $p_d : S^1 \to S^1$ given by $p_d(x) = dx \bmod 1$. For $d = 1$ we obtain the identity map.

**Example: the projective plane.** This theorem gives a convenient way to *compute* fundamental groups. For example, we have

$$\pi_1(\mathbb{RP}^2) \cong \mathbb{Z}/2.$$

**Examples: a bouquet of two circles.** See Figure 7 for examples of coverings of the $S^1 \vee S^1$, one with Galois group $S_3$, the other of degree 3 and irregular.

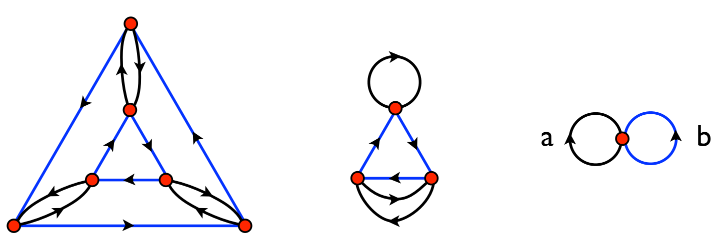



Figure 7. Regular and irregular coverings of the bouquet of two circles.

**Universal covers.** We now wish to show that all subgroups arise from covering spaces. In particular we must show that the trivial group arises.

If $\pi_1(E)$ is trivial, we say $E$ is the *universal cover* of $B$. By the preceding considerations, it is unique up to isomorphism over $B$. This cover is 'universal' in the sense that it lies above any other (connected) covering space $E'/B$.

**Path space.** Given a pointed topological space $(B, b)$, we let $\mathcal{P}(B, b)$ denote the *path space* of all continuous functions $f : [0, 1] \to B$ such that $f(0) = b$. The path space has a natural basepoint, namely the constant map to $b$. If $B$ is path–connected then $\mathcal{P}(B, b)$ is connected.

Suppose for the moment that $B$ is a metric space; then the path space carries a natural topology.

**Theorem 13.6** *Path space is contractible.*

**Proof.** The maps $F_s : \mathcal{P}(B, b) \to \mathcal{P}(B, b)$, given by $F_s(f) = f_s$ where $f_s(t) = f(st)$ for $s \in [0, 1]$, give a deformation retract of the path space to its basepoint. ∎

**Loop spaces, etc.** We will show a *quotient* of the path space $\mathcal{P}(B, b)$ is its universal cover. However it is important to note that in general, $\widetilde{B}$ is *not* contractible (e.g. when $B = S^n$, $n > 1$). Thus the quotient operation introduces *new topology* into the path space. From a more sophisticated perspective, we have a *fibration*

$$\Omega(B, b) \to \mathcal{P}(B, b) \to B$$

with fiber the *loop space* of $B$. The homotopy theory of $B$ persists, in an shifted form; namely $\pi_{i+1}(B) \cong \pi_i(\Omega(B))$. For example, the components of the loop space correspond to the elements of $\pi_1(B)$, since a loop in once class cannot be connected to a loop in another class.

**Theorem 13.7** *If $B$ is locally simply–connected, then $B$ has a universal cover.*

**Proof.** Fix a basepoint $b \in B$. Let $\widetilde{B}$ be the set of *homotopy classes* of paths $f : ([0, 1], 0) \to (B, b)$. The constant path $c_b$ provides a natural basepoint $[c_b] \in \widetilde{B}$. There is a natural map

$$p : \widetilde{B} \to B \quad \text{given by} \quad p(f) = f(1),$$

and satisfying $p(\widetilde{b}) = b$. For any open neighborhood $U$ of $f(1)$, we can consider the collection $U(f)$ of paths $[f * g] \in \mathcal{B}$ such that $g : [0,1] \to U$. The sets $U(f)$ are a basis for a topology on $\widetilde{B}$. Note that $\widetilde{B}$ is connected, because $\mathcal{P}(B,b)$ is connected.

We claim that $p : (\widetilde{B}, \widetilde{b}) \to (B, b)$ gives the universal cover of $B$.

We should first check that $p$ is a covering map. Given $x \in B$, choose a simply-connected neighborhood $U$ of $x$. Then for any $f_i \in \widetilde{B}$ with $p(f_i) = x$, we have a natural map

$$p_i : U \to U(f_i)$$

given by $p_i(y) = [f * h]$, where $h$ is a path from $x$ to $y$ in $U$. This map is well–defined because $U$ is simply–connected, and it satisfies $p \circ p_i(y) = y$. Thus $p : \widetilde{B} \to B$ covers $B$ evenly.

Next, note $G = \pi_1(B, b)$ acts naturally on $\widetilde{B}$ by $[g] \cdot [f] = [g * f]$. This action is *transitive on fibers*: if $p(f_1) = p(f_2)$, then $g = f_1 * f_2^{-1}$ represents an element of $G$, and $[g] \cdot [f_2] = [f_1]$. In fact $G$ is naturally identified with $\mathrm{Gal}(\widetilde{B}/b)$. Thus $\widetilde{B}$ is the universal cover of $B$.

Alternatively, one can check directly that $\widetilde{B}$ is simply–connected. Let $F : [0,1] \to \widetilde{B}$ be a loop based at $\widetilde{b}$. Then $f(t) = p(F(t))$ is a loop based at $b$ in $B$. The assumption that $F(1) = \widetilde{b}$ *means* that $[f] = \mathrm{id}$ in $\pi_1(B, b)$. Then a deformation of $f$ to the constant map gives a deformation of $F$ to $c_b$, and thus $[F] = \mathrm{id}$ in $\pi_1(\widetilde{B}, \widetilde{b})$. ■

Note: The universal cover can be constructed in a slightly more general setting, namely that of a semilocally simply–connected space; see Munkres.

**Theorem 13.8** *Every subgroup $H \subset G = \pi_1(B, b)$ arises from a covering space $E$, with*

$$\pi_1(E, e) \cong H \cong \mathrm{Gal}(\widetilde{B}/E).$$

*If $H \subset G$ is a normal subgroup, then $E/B$ is a normal covering and*

$$\mathrm{Gal}(E/B) \cong G/H.$$

**Proof.** Letting $E = H \backslash \widetilde{B}$, it is easy to check that $H = H_E = \pi_1(E)$. If $H$ is normal, then $G = \mathrm{Gal}(\widetilde{B}/B)$ acts by deck transformations on $E$ by

$$g \cdot (Hx) = (gH) \cdot x = H \cdot (gx),$$

with $H$ acting trivially. This action is transitive on fibers since $G$ acts transitively on the fibers of $\widetilde{B}/B$. ■

**Corollary 13.9** *Covering spaces of $(B,b)$ with deck group $K$, up to isomorphism over $B$, correspond to surjective homomorphisms $\phi : \pi_1(B,b) \to K$.*

**Proof.** Take the covering $E/B$ defined by the normal subgroup $\mathrm{Ker}(\phi)$. ■

The description of the covering space $E$ corresponding to $H \subset G = \pi_1(B,b)$ is summarized by the following diagram:

$$\begin{tikzcd}
\widetilde{B} \arrow[dd, "G"'] \arrow[dr, "H"] & \\
& E \arrow[dl, "G/H\ \text{(if $H$ is normal)}"] \quad = H\backslash G, \quad \pi_1(E) = H \\
B. &
\end{tikzcd}$$

**Comparison with Galois theory.** Let $K/k$ be a finite field extension. Then $G = \mathrm{Gal}(K/k)$ is the group of $k$-linear automorphisms of the field $K$. For any subgroup $H \subset G$ we let $K^H \subset K$ denote the subfield fixed pointwise by all elements in $H$. The field extension is *Galois* if $K^G = k$. In this case the subfields between $K$ and $k$ correspond exactly to the subgroups $H \subset G$, and $K^H/k$ is Galois iff $H$ is normal in $G$.

Thus $K/k$ is a good analogue of a normal, or Galois, covering space. The concept of a universal cover is more subtle, however. It corresponds to the algebraic closure $\overline{k}$ of $k$, but the analogy is not perfect. In fact $\mathrm{Gal}(\overline{k}/k)$ is a profinite group, and it is analogous to the profinite completion of $\pi_1(B,b)$.

As an example, let $k = \mathbb{Q}$ and let $K = \mathbb{Q}(\omega, 2^{1/3}) \subset \mathbb{C}$, where $\omega^3 = 1$. Then $\deg(K/k) = 6$ and in fact $\mathrm{Gal}(K/k) \cong S_3$. Indeed, $K$ is the splitting field of the equation $X^3 - 2$. This equation has 3 roots, $2^{1/3}\omega^i$ with $i = 0, 1, 2$. Clearly $K$ is the smallest field containing these 3 roots, and any element of $\mathrm{Gal}(K/k)$ is determined by its action on these 3 roots. In other words, we have an injective map $\mathrm{Gal}(K/k) \to S_3$. But there is an obvious element of order 2 in $\mathrm{Gal}(K/k)$, namely $z \mapsto \overline{z}$. Thus $\mathrm{Gal}(K/k) = S_3$.

The fields between $K$ and $k$ correspond bijectively, then to subgroup of $S_3$. For example, $\mathbb{Q}(2^{1/3}) = L$ is degree 3 over $\mathbb{Q}$, but *not Galois.* There is only one root to $X^3 - 2$ in this field! So it cannot be moved. The field $L$ is the fixed field of complex conjugation. On the other hand, $\mathbb{Q}(\omega) = M$ is degree 2 over $\mathbb{Q}$, and hence it has Galois group $\mathbb{Z}/2$, given by complex conjugation. The field $M$ is the fixed field of an element of order 3 in $\mathrm{Gal}(K/k)$, which cyclically permutes the cube roots of two act (therefore) acts linearly over $\mathbb{Q}(\omega)$.

$$
\begin{tikzcd}
 & \mathbb{Q}(\omega, 2^{1/3}) \arrow[dl, dash, "\mathbb{Z}/2"'] \arrow[dr, dash, "\mathbb{Z}/3"] \arrow[dd, dash, "S_3"'] & \\
\mathbb{Q}(2^{1/3}) \arrow[dr, dash] & & \mathbb{Q}(\omega) \arrow[dl, dash, "\mathbb{Z}/2"'] \\
 & \mathbb{Q} &
\end{tikzcd}
$$

The diagram above summarize the relationships between these fields.

# 14 Free groups and graphs

In this section we describe the fundamental groups of 1-dimensional simplicial complexes (graphs), and begin a study of groups and their presentations by discussing free products.

**Free products.** There are many ways to discuss free groups. Let us beginning with the functorial definition via universal properties. Let $G_1 * G_2$ be a group equipped with distinguished subgroups $G_1$ and $G_2$. We say $G_1 * G_2$ is the *free product* of $G_1$ and $G_2$ if given maps $G_i \to G$, $i = 1, 2$, there is a unique map $G_1 * G_2 \to G$ making the diagram

$$
\begin{tikzcd}
G_1 \arrow[d] \arrow[dr] & \\
G_1 * G_2 \arrow[r] & G \\
G_2 \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

commute.

It can be shown (see below) that free products always exist. But even without knowing this, uniqueness can be verified by the properties above.

**Bouquets of circles and Cayley graphs.** Let $S^1 \vee S^1$ be a bouquet of two oriented circles, labeled $a$ and $b$, joined at the basepoint $*$. We can think of $a$ and $b$ as elements of $\pi_1(S^1 \vee S^1, *)$. A basic and intuitive fact is:

**Theorem 14.1** *The loops $a$ and $b$ generate $\pi_1(S^1 \vee S^1, *)$.*

**Cayley graphs.** Now let $G$ be a group generated by two elements, also called $a$ and $b$. The *Cayley graph* $\Gamma$ of $G$ is the oriented 1-complex defined by taking the elements $x \in G$ as vertices, and joining $x$ to $y$ with an edge whenever $x = ya$ or $x = yb$. We oriented each such edge from $x$ to $y$ and label it by $a$ or $b$. Then every vertex has degree 4, with 2 incoming edges and 2 outgoing edges.

We define a map

$$p : \Gamma \to S^1 \vee S^1$$

by sending all the vertices of the Cayley graph to the basepoint $*$, and sending the $a$-edges to the $a$-loop and the $b$-edges to the $b$-loop, respecting orientations. Then clearly $p$ is a covering map.

Note that $p$ carries the *same information* as the decoration of the edges of $\Gamma$. That is, $p$ determines this edge decorations, and the decorations determine $p$.

**Deck transformations.** In particular, $\mathrm{Gal}(\Gamma/S^1 \vee S^1)$ is the same as the group of automorphisms of $\Gamma$ as a graph with decorated edges. Moreover, $G$ acts on $\Gamma$ by $x \mapsto gx$. Since right and left multiplication commute, this action preserves the edge labelings. Conversely, an automorphism $f$ of $\Gamma$ preserving decorations is unique determined once we know $f(e)$; if $f(e) = g$, then $f(x) = xg$, as can be seen by writing $x$ as a product of the generators, tracing out edges from $e$ to $g$, and then taking their image.

It follows that:

$p : (\Gamma, e) \to (S^1 \vee S^1, *)$ *is a Galois covering space with deck group $G$.*

More precisely, the natural homomorphism $\phi : \pi_1(S^1 \vee S^1, *) \to \mathrm{Gal}(\Gamma/X)$ is given by $\phi(a) = a$ and $\phi(b) = b$. This homomorphism is unique because $a$ and $b$ generate $\pi_1(S^1 \vee S^1, *)$. In other words, the universal property required for a free product is satisfied. We have therefore shown:

**Theorem 14.2** *The fundamental group of $S^1 \vee S^1$ is isomorphic to the free product $F_2 = \mathbb{Z} * \mathbb{Z}$.*

By the same reasoning, $\pi_1(\vee^n S^1) \cong F_n$ and more generally we have:

**Theorem 14.3** *The fundamental group of any finite connected graph $\Gamma$ is free. In fact $\pi_1(\Gamma) \cong F_n$ where $n = |E(\Gamma)| - |V(\Gamma)| + 1$.*

**Proof.** Choose a maximal tree $T \subset \Gamma$; then $\Gamma/T$ is a bouquet of $n$ circles. ∎

**Remark: Euler characteristic.** The quantity $V - E$ for a graph, $V - E + F$ for a surface, and more generally the *Euler characteristic* $\chi(X) = \sum_K (-1)^{\dim(K)}$ for a simplicial complex, are all homotopy invariants. The case of a tree is the simplest example.

Here is a related result for *infinite* graphs.

**Theorem 14.4** *For an infinite, connected graph $Y$, the following are equivalent:*

1. *$\pi_1(Y)$ is finitely–generated.*
2. *There is a finite subgraph $K$ such that $\pi_1(K)$ is isomorphic to $\pi_1(Y)$.*
3. *There is a finite subgraph $K$ such that every edge outside $K$ disconnected $Y$.*
4. *Every component of $Y - K$ is a tree rooted in $K$.*

**Proof.** Suppose $Y$ is finitely–generated. Then by considering the support of the generating loops, we obtain a finite subgraph $K$ such that $\pi_1(K)$ maps *onto* $\pi_1(Y)$. Now suppose there is an edge $E$ in $Y - K$ that does not disconnect $Y$. Then we can collapse $Y - E$ to a point and obtain a map $p : Y \to S^1$, surjective on $\pi_1$. But $p(K)$ is a point, so $\pi_1(K)$ did not map surjectively to $\pi_1(Y)$. Thus (1) implies (2) $\pi_1(Y) \cong \pi_1(K)$ as well as (3). But (3) implies (4) and (4) implies $Y$ deformation retracts to $K$, which implies (1) and (2). ∎

**Subgroups of free groups.** Here is a famous theorem that can be proved using graph theory.

**Theorem 14.5** *Any finitely generated subgroup $H$ of a free group $F_n$ is free; that is, $H \cong F_m$ for some $m$.*

**Proof.** Let $X$ be a bouquet of $n$ circles; then $\pi_1(X) \cong F_n$ and so $H$ determine a covering space $Y \to X$ with $\pi_1(Y) \cong H$. Clearly $Y$ is a graph. Since $\pi_1(Y)$ is finitely generated, there is a finite subgraph such that $\pi_1(K) \cong \pi_1(Y)$, and we have seen that $\pi_1(K) \cong F_m$ for some $m$. ∎

**Examples and counterexamples.** Any degree 2 cover of $S^1 \vee S^1$ gives a copy of the $F_3$ inside $F_2$. A concrete example is given by $\langle a, b^2, bab\rangle$ inside $F_2 = \langle a, b\rangle$.

The kernel of the map $F_2 \to \mathbb{Z}$ sending $b$ to zero is *infinitely generated*.

**Free products of general groups.** Let $G_i$ be a collection of groups. As sets, we regard all the groups as having the same identity element $e$; otherwise the groups are disjoint. The free product $G = * G_i$ can be defined by a universal property as before. But our goal now is to *explicitly construct* this group.

First, a *word* $w$ is a finite sequence of elements $w = (a_1, a_2, \ldots, a_n)$ with $a_j \in \bigcup G_i$. Let $W$ denote the set of all words (including the word of length zero). We define a product on $W$ by concatenation; that is,

$$w * v = (a_1, \ldots, a_n) * (b_1, \ldots, b_m) = (a_1, \ldots, a_n, b_1, \ldots, b_m).$$

A word is *reduced* if $a_j$ and $a_{j+1}$ belong to different $G_i$'s for all $j$. As a set, $G$ is the collection of all reduced words.

Second, given a word $w$ that may not be reduced, a *shortening* of $w$ is any word of the form

$$w' = (a_1, \ldots, a_j a_{j+1}, \ldots, a_n),$$

where $a_j$ and $a_{j+1}$ lie in the same group $G_i$. By iterated shortening we eventually arrive at a reduced word.

**Theorem 14.6** *All shortening lead to the same reduced word $r(w)$.*

**Proof.** The proof is by induction on the length of $w$, being clear if the length of $w$ is 1. Now let $w'$ and $w''$ be two different shortenings of $w$, working on the letters with indices $j < k$. In other words, $w'$ is obtained by replacing $(a_j, a_{j+1})$ with $(a_j a_{j+1})$, and $w''$ is obtained by replacing $(a_k, a_{k+1})$ with $(a_k a_{k+1})$.

If the letters are far enough apart, we can then simplify the $k$-pair in $w'$ and the $j$-pair in $w''$ to obtain a common simplification $w'''$ of both. Thus there exists a chain of simplifications of $w'$ and $w''$ leading to the same reduced word. By induction, all simplifications of $w'$ and $w''$ lead to the same word, so we are done.

It may however happen that $k = j + 1$. But then we have a triple of letters $(a_j, a_{j+1}, a_{j+2})$ all lying in the same group $G_i$. Thus we can simplify $w'$ and $w''$ to obtain a word $w'''$ where this triple of letters is replaced by $a_j a_{j+1} a_{j+2}$. Again we are done by induction. ∎

**Corollary 14.7** *We have* $r(v*w) = r(r(v)*w) = r(v*r(w)) = r(r(v)*r(w))$.

We now define the product operation on $G$ by $vw = r(v * w)$. Because of the preceding theorem, it is now easy to verify the group axioms. For example, the product is associative because

$$(uv)w = r(r(u * v) * w) = r(u * v * w) = r(u * r(v * w)) = u(vw).$$

**Universality.** We can now verify that $G$ is an explicit model for the free product of the groups $G_i$. Note: the only reason $W$ itself doesn't work is that $W$ has no inverses!

**Theorem 14.8** *Any collection of homomorphisms* $\phi_i : G_i \to H$ *extends to a unique homomorphism* $\phi : G \to H$.

**Proof.** Define a map $\phi : \bigcup G_i \to H$ by taking the union of the $\phi_i$. Then we have a unique extension $\phi : W \to H$ defined by $\phi(a_1, \ldots, a_n) = \phi(a_1) \cdots \phi(a_n) \in H$. By definition, we have

$$\phi(v * w) = \phi(v)\phi(w).$$

Now it is easy to check that $\phi(w) = \phi(w')$ whenever $w'$ is a simplification of $w$. Therefore $\phi(r(w)) = \phi(w)$. Restricting to the set $G \subset W$ consisting of reduced words, we find

$$\phi(vw) = \phi(r(v * w)) = \phi(v * w) = \phi(v)\phi(w),$$

so $\phi$ is a homomorphism. ∎

**Reflection groups.**

1. Let $G$ be the group generated by reflections $a$ and $b$ in two lines in $\mathbb{R}^2$ meeting at $\pi/n$ degrees. Then $a^2 = b^2 = \mathrm{id}$, and $ab$ is rotation by $2\pi/n$ degrees, so $(ab)^n = \mathrm{id}$. The Cayley graph is $G$ thus looks like a regular $n$-gon.

2. Now let $n \to \infty$. Then $G$ 'converges to' the group $\mathbb{Z}/2 * \mathbb{Z}/2$ generated by $a$ and $b$ with $a^2 = b^2 = \mathrm{id}$ as the only relation; that is, we have

$$G \cong \mathbb{Z}/2 * \mathbb{Z}/2.$$

Any word in this group is simply an alternating sequence of $a$'s and $b$'s, such as $g = abababa$. The Cayley graph is a straight line, and $G$ can be visualized geometrically as the group generated by reflections in two *parallel* lines in $\mathbb{R}^2$. In other words we can take $a(x) = -x$ and $b(x) = 2 - x$ acting on $\mathbb{R}$.

We can also consider $G$ as an infinite dihedral group with $f = a$ and $r = ab$. Then $r^{-1} = ba$, and hence

$$fr = a(ab) = b = (ba)a = r^{-1}f.$$

3. $G = \mathbb{Z}/2 * \mathbb{Z}/2 * \mathbb{Z}/2$. Here we have $3^n$ words of length $n$. This work can be realized as isometries of the hyperbolic plane, generated by reflections in the sides of an ideal triangle. (Other triangles will do as well!)

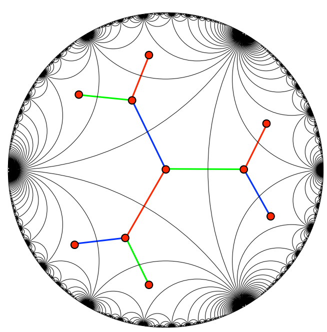


Figure 8. Hyperbolic geometry and $\mathbb{Z}/2 * \mathbb{Z}/2 * \mathbb{Z}/2$.

In the upper halfplane model for $\mathbb{H}$, we can think of $G$ as the group generated by reflections in the 'lines' (geodesics) $[1, \infty]$, $[-1, \infty]$ and $[-1, 1]$. (The latter is the unit semicircle centered at $z = 0$.) The generators can then be given directed by the transformations $a(z) = 2 - \overline{z}$, $b(z) = -2 - \overline{z}$, and $c(z) = 1/\overline{z}$.

4. What does the orientation–preserving subgroup $G_0$ look like? In the free group on involutions $a, b, c$, a typical word looks like *abacbcbac*. We just need to be careful never to use the same letter twice in a row. Thus the words of even length for a subgroup generated by $ab$, $bc$ and $ac$. But $(ab)(bc) = (ac)$, so in fact the words of even length give the *free group* on 2 generators.

This can be seen intuitively by noting that $\mathbb{H}/G$ is a triply–punctured sphere.

# 15 Group presentations, amalgamation and gluing

In this section we discuss how the fundamental group of $X \cup Y$ is related to the fundamental groups of $X$ and $Y$. As one might expect from the discussion of the previous section, if $X$ and $Y$ are reasonable spaces (with basepoints), then

$$\pi_1(X \vee Y) \cong \pi_1(X) * \pi_1(Y).$$

Here we wish to allow for $X$ and $Y$ to overlap in a more complicated way. The general case allows one to compute a presentation for the fundamental group of any simplicial complex. For example, we will show:

**Theorem 15.1** *Given any finitely presented group $G$, there exists a finite 2-complex $(X, x)$ such that $\pi_1(X, x) \cong G$.*

**Killing group elements.** Let $G$ be a group. We begin with a general discussion of killing an element $r \in G$. This means we want to construct a new group $G'$ where $r = \mathrm{id}$, but otherwise the group is as large as possible.

The solution is obtained using $N(r)$, the smallest *normal* subgroup containing $r$. This normal subgroup exists by general considerations; it can be constructed from the 'outside' by intersecting all normal subgroups containing $r$. On the other hand, $N$ can also be constructed from the inside: it is generated by the elements $grg^{-1}$ where $g$ ranges over *all* elements of $G$. Note that it general *not* to let $g$ range over the generators of $G$. Even with $G$ is finitely generated, $N(r)$ may be *infinitely generated*. See below.

To kill $r$ we simply set

$$G' = G/N(r).$$

Note that if $\phi : G \to H$ and $\phi(r) = \text{id}$, then we *must* have $N(r) \subset \text{Ker}(\phi)$, since the kernel of $\phi$ is normal. This shows $G'$ is the *largest* quotient of $G$ with $r = \text{id}$.

**Group presentations.** Let $a_1, \ldots, a_n$ be generators for a free group $F_n$. With these chosen generators we write

$$F_n = \langle a_1, \ldots, a_n \rangle.$$

This group is as free as possible, in the sense that any map $\phi : \{a_1, \ldots, a_n\} \to H$ extends to a homomorphism $\phi : F_n \to H$.

We now wish to describe a smaller group

$$G = \langle a_1, \ldots, a_n | r_1, \ldots, r_m \rangle,$$

where we have imposed certain relations on the generators. The relations $r_i$ are words in the generators $a_i$. We can assume they are reduced words, so we may take $r_i \in F_n$. Then $G$ is the largest group where these relations *hold*. In other words, it should have the property that for any choice of elements $\phi(a_i) \in H$ such that $\phi(r_j) = 0$ for all $j$, we get a unique homomorphism $\phi : G \to H$.

The group $G$ can be constructed formally by taking

$$G = F_n / N(r_1, \ldots, r_m),$$

where $N(r_1, \ldots, r_m)$ is the smallest normal subgroup of $F_n$ containing the given relations.

It is then clear that $G$ has the desired universal property.

**One relator can give an infinitely–generated kernel.** For a simple example, let $G = \langle a, b | b \rangle$. Then $N(b)$ is generated by $a^i b a^{-i}$ for all $i$, and as we have seen in the study of graphs the kernel of the map $F_2 \to G \cong \mathbb{Z}$ really is infinitely generated in this case.

**What is $G$?** If we are given a group $H$ with generators $a_1, \ldots, a_n$ which satisfy relations $r_1, \ldots, r_m$, then of course we get a surjective map

$$\phi : G = \langle a_i | r_j \rangle \to H.$$

Some care is required, however, to check that this map is an isomorphism. When it is, we say $H$ is *finitely presented*, and the given $(a_i | r_j)$ are a *presentation* for $H$.

**Examples.**

1. We have $G = \langle a \rangle \cong \mathbb{Z}$.
2. We have $G = \langle a, b \ : \ ab = ba \rangle \cong \mathbb{Z}^2$. (An expression of the form $w_1 = w_2$ means, of course, $w_1 w_2^{-1}$ is a relator.) To see this, we first note that $G$ is *abelian*. Thus every element of $G$ can be written in the form $a^i b^j$, and the map to $\mathbb{Z}^2$ is given by recording the exponents $(i, j)$. The only element that maps to zero is $a^0 b^0 = \text{id}$, so $G \cong \mathbb{Z}^2$.
3. We have $G = \langle r, f \ : \ r^n = f^2 = id, rf = fr^{-1} \rangle \cong D_{2n}$. Using the last relation, we can write every element of $G$ in the form $r^i f^j$ for some $i, j$. Then the first two relations show we can take $0 \leq i < n$ and $0 \leq j < 2$. Thus $G$ has at most $2n$ elements, so the map to $D_{2n}$ is an isomorphism.
4. The symmetric group $S_{n+1}$ is generated by transpositions $\sigma_i = (i, i+1)$ subject to the relations:

   $\sigma_i^2 = e$;
   $\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$; and
   $\sigma_i \sigma_j = \sigma_j \sigma_i$ if $|i - j| > 1$.

   This is nontrivial to verify. The main point is to show that if a complicated word in the $(\sigma_i)$ maps to the identity in the free group, then it can be simplified to the identity using the relations above. The point of these relations is that they allow one to change or simplify crossings in a braid diagram. (Remark: if we drop the condition that $\sigma_i = e$, we get the braid group!)

**Free product with amalgamation.** Let $\psi_i : A \to G_i$, $i = 1, 2$ be a pair of homomorphisms. We define $G_1 *_A G_2$ by the following universal property: Given $\phi_i : G_i \to G$, $i = 1, 2$ such that $\phi_1 \circ \psi_1 = \phi_2 \circ \psi_2$ on $A$, there is a unique homomorphism $\phi : G_1 *_A G_2 \to G$ making the diagram:

$$\begin{tikzcd}
 & G_1 \arrow[d] \arrow[dr] & \\
A \arrow[ur] \arrow[dr] & G_1 *_A G_2 \arrow[r] & G \\
 & G_2 \arrow[u] \arrow[ur] &
\end{tikzcd}$$

commute.

Note that we now have natural *maps* $G_i \to G_1 *_A G_2$, but these maps need not be injections, so $G_i$ can no longer be regarded as *subgroups* of $G_1 *_A G_2$. Indeed, if $A$ maps isomorphically to $G_1$ but trivially to $G_2$, then $G_1$ maps trivially to $G_1 *_A G_2$.

As another example, if $A = \mathbb{Z}$ maps to $r \in G_1$ and $G_2$ is trivial, then we are just killing $r$:

$$G_1 *_A G_2 \cong G_1/N(r).$$

This process arises naturally when we attach a 2-cell to a complex.

**Theorem 15.2** *A presentation for the free product with amalgamation $G_1 *_A G_2$ is obtained from a presentation for $G_1 * G_2$ by adjoining the relations $\psi_1(a_i) = \psi_2(a_i)$ as $a_i$ ranges over a set of generators for $A$.*

**Proof.** Let $(a_1, \ldots, a_n)$ be generators for $A$, and let $H = G_1 * G_2/N$, where $N$ is the smallest normal subgroup containing the set $\psi_1(a_i)\psi_2(a_i)^{-1}$. Now suppose maps $\phi_i : G_i \to G$ are given making the diagram above commute. Then $\phi_1 \cup \phi_2$ extends to a unique map $\Phi : G_1 * G_2 \to G$. Compatibility with the maps $\psi_i : A \to G_i$ means $N \subset \operatorname{Ker}(\Phi)$. Thus $\Phi$ descends to a unique map $\phi : H \to G$, and thus $H$ satisfies the universal property required for $G_1 *_A G_2$. ∎

**Gluing: Seifert–van Kampfen.** Let $X$ be a reasonable space — connected, locally path–connected and locally simply–connected. Let $X = U_1 \cup U_2$ be a union of two open, connected sets, such that $V = U_1 \cap U_2$ is also connected. Let $x \in V$ be a basepoint. Then we have natural maps

$$\begin{tikzcd} & \pi_1(U_1, x) \arrow[dr] & \\ \pi_1(V, x) \arrow[ur, "i"] \arrow[dr, "j"'] & & \pi_1(X, x) \\ & \pi_1(U_2, x) \arrow[ur] & \end{tikzcd}$$

The following result says $\pi_1(X, x)$ is built from these pieces as freely as possible, subject to the obvious constraints.

**Theorem 15.3 (Seifert–van Kampfen)** *The fundamental group of $X$ satisfies*

$$\pi_1(X) \cong \pi_1(U_1) *_{\pi_1(V)} \pi_1(U_2).$$

**Example: simple connectivity.** If $U_1$, $U_2$ and $U_1 \cap U_2$ are all simply-connected, then so is $X = U_1 \cup U_2$. This gives, for example, an inductive proof that $\pi_1(S^n)$ is trivial for $n \geq 2$.

**Joins.** As a second basic example, suppose $X_1$ and $X_2$ are finite simplicial complexes with basepoints. Then we can slightly thicken their basepoints to obtain an expression

$$X = X_1 \vee X_2 = U_1 \cup U_2$$

with $U_1 \cap U_2$ contractible, and the inclusion $X_i \subset U_i$ a homotopy equivalence for $i = 1, 2$. Then we have

$$\pi_1(X_1 \vee X_2) \cong \pi_1(X_1) * \pi_1(X_2).$$

**Disconnected covers.** We will prove Theorem 15.3 following an idea of Grothendieck.

When classifying regular covers, we did not make use of all the homomorphism from $\pi_1(B, b)$ to $G$, only the surjective ones. This is because we wanted $E$ to be connected. Following Grothendieck, we now drop this assumption.

Given a group $G$, a *$G$-cover of $B$* is a *possibly disconnected* covering space

$$p : (E, e) \to (B, b),$$

equipped with a *free* action of $G$ on $E$ by deck transformations, such that $G$ acts transitively on the fibers of $p$. For example, $(E, e) = (G, \mathrm{id})$ can be regarded as a $G$-cover of a point. Note that $G$ need not be the *full* group of deck transformations; e.g. the full deck group of a degree $d$ covering of a point is $S_d$, but we only want the free action of a group of order $d$. Note also that in the case of a *connected* cover, the deck group always acts freely, but here we *impose* this property on $G$.

The notion of isomorphism between $G$-covers $E$ and $E'$ is the obvious one: it is given by a homeomorphism $h : (E, e) \to (E', e')$ respecting the $G$-action and projection to the base.

**Theorem 15.4** *There is a natural bijection between $G$-covers of $(B, b)$, up to isomorphism over $(B, b)$, and the elements $\phi \in \mathrm{Hom}(\pi_1(B, b), G)$.*

**Idea of the proof.** Given a $G$-covering $p : (E, e) \to (B, b)$, let $E_0$ denote the component of $E$ containing the basepoint $e$, and let $H \subset G$ denote the subgroup that sends $E_0$ to itself. Since $G$ acts transitively on fibers, we see

$E_0/B$ is a *connected* Galois covering with deck group $H$, and thus we have a natural homomorphism

$$\phi : \pi_1(B, b) \to \mathrm{Gal}(E_0/B) \cong H \subset G.$$

Conversely, given $\phi : \pi_1(B, b) \to G$, we obtain an action of $\pi_1(B, b)$ on $(\widetilde{B}, \widetilde{b}) \times G$ by

$$\gamma \cdot (x, g) = (\gamma \cdot x, g\phi(\gamma)^{-1}).$$

We can then form the quotient space

$$E = \widetilde{B} \times_{\pi_1(B)} G,$$

with basepoint $e$ given by the image of $(\widetilde{b}, \mathrm{id})$. Now $G$ acts on $E$ by

$$g' \cdot [(x, g)] = [(x, g'g)],$$

respecting equivalence classes since right and left multiplication on $G$ commute. The projection $p : \widetilde{B} \to B$ then provides a natural projection

$$[(x, g)] \mapsto p(x) : (E, e) \to (B, b).$$

It is readily verified that we have constructed a $G$-cover. Thus we have procedures for going from $G$-covers to homomorphisms $\phi : \pi_1(B, b) \to G$ and back again. It is readily verified that the result is a bijection. ■

**Example of a somewhat irregular *G*-cover.** Since $G$ acts transitively on the fibers of a $G$-cover, it is tempting to think of all such coverings as 'normal'. However different behavior arises depending on whether or not the *image* of $\pi_1(B)$ in $G$ is normal or not.

As an example, let $E \to S^1$ be the covering space corresponding to the map

$$\phi : \pi_1(S^1) \cong \mathbb{Z} \to \mathbb{Z}/2 = H \subset G = S_3.$$

This map sends the generator of $\mathbb{Z}$ to a transposition $\tau \in S_3$. The subgroup $H = \mathrm{Im}(\phi)$ is not normal in this example.

The covering space $E$ is evidently isomorphic to $S^1 \times G/H$, i.e. it is the disjoint union of 3 circles, each mapping by degree 2 to the base. The transposition $\tau \in G$ acts on the component $E_0$ containing the basepoint by its unique nontrivial deck transformation: $E_0/B$ is a Galois covering with deck group $\langle \tau | E_0 \rangle$. However $\tau$ *swaps* the other two components of $E$. Indeed, the components of $E$ can be identified with $G/H$, a set with 3 points, and the action of $G$ on components is just the usual action of $S_3$ by permutations.

**Proof of Theorem 15.3.** Let $G_1, G_2$ and $A$ be $\pi_1(U_1)$, $\pi_1(U_2)$ and $\pi_1(U_1 \cap U_2)$. Then (for a basepoint in $U_1 \cap U_2$) we have natural maps $\psi_i : A \to G_i$.

Now suppose we are given maps $\phi_i : G_i \to G$ which are compatible on $A$; that is, such that $\phi_1 \circ \psi_1 = \phi_2 \circ \psi_2$. Then we have $G$-coverings of $U_1$ and $U_2$ which restrict to the *same* $G$-covering of $U_1 \cap U_2$. Thus they can be patched together to obtain a $G$-covering of $X = U_1 \cap U_2$, and hence a homomorphism $\phi : \pi_1(X) \to G$. This homomorphism is unique since $G_1$ and $G_2$ generate $\pi_1(X)$. Thus $\pi_1(X)$ has the universal property which characterizes $G_1 *_A G_2$. ∎

**Examples.** Let $\Sigma_g$ denote a closed surface of genus $g$.

- Think of a torus $X = \Sigma_1$ as a square $S$ with opposite sides identified. Remove a subsquare $S'$; then the result $U$ is a torus with boundary, and it retracts onto $S^1 \vee S^1$, so $\pi_1(U) \cong \langle a, b \rangle \cong F_2$. Thicken $S'$ slightly to obtain an open disk $V$. Then $U \cap V$ is a neighborhood of a square, with $\pi_1(U \cap V) = \langle c \rangle \cong \mathbb{Z}$. The inclusion to $V$ kills $c$, while the inclusion to $U$ gives $c = [a, b]$. Thus
$$\pi_1(\Sigma_1) \cong \langle a, b, c \;:\; [a, b] = c = \mathrm{id} \rangle \cong \mathbb{Z}^2.$$

- A surface of genus two can be obtained by gluing together two copies of $U$ as above, with $\pi_1(U_i) \cong \langle a_i, b_i \rangle \cong F_2$. Thus:
$$\pi_1(\Sigma_2) \cong \langle a_1, a_2, b_1, b_2 \;:\; [a_1, b_1] = [a_2, b_2] \rangle.$$

- The dunce cap is simply connected: we have
$$\pi_1(D) \cong \langle a \;:\; aaa^{-1} = \mathrm{id} \rangle.$$

- The Klein bottle satisfies
$$\pi_1(K) \cong \langle a, b \;:\; ab = ba^{-1} \rangle.$$
If we added the relation $b^2 = 1$, it would be the infinite dihedral group $D_\infty$.

Thus $K$ admits a Galois covering with Galois group $D_\infty$. Explicitly, there is an action of $D_\infty$ on the infinite cylinder $C = S^1 \times \mathbb{R}$ given by

$$f(x,t) = (-x,-t) \quad \text{and} \quad r(x,t) = (x,t+1),$$

and the quotient is the Klein bottle. We can also observe that $D_\infty$ has a normal subgroup $N = \langle r \rangle \cong \mathbb{Z}$, that $C/N \cong S^1 \times \mathbb{R}/\mathbb{Z}$ is a torus, and that $D_\infty/N \cong \mathbb{Z}/2$ acts on $C/N$ by the composition of the translation $(x,t) \mapsto (-x,t)$ with the reflection $(x,t) \mapsto (x,-t)$. (Note that $-x$ is the antipodal map on $S^1$, while $-t$ is the reflection on $\mathbb{R}/\mathbb{Z}$.)

- The real projective plane, presented as a quotient of the square $S$, satisfies
$$\pi_1(P) \cong \langle a \ : \ a^2 = 1 \rangle.$$
It is tempting to think the presentation is $G = \langle a, b \ : \ (ab)^2 = 1 \rangle$ — a much bigger group! (For example, $G$ maps onto $\mathbb{Z}$ by $a \mapsto 1$, $b \mapsto -1$.) But in fact $\partial S$ modulo gluing relations is not two circles, but one: there are 2 vertices (and 2 edges) in the quotient.

**Theorem 15.5** *Given any finitely presented group $G$, there exists a finite 2-complex $K$ such that $\pi_1(K) \cong G$.*

**Proof.** Start with a bouquet of circles with a loop for each generator of $G$, and then attach an $n$-gon for each relator of length $n$. ∎

**The knot group.** This is the group of flight plans for airplanes leaving from a base located outside the knot and flying around it.

Examples: For the unknot, $G(K) = \mathbb{Z}$. For the unlink on two components, $G(L) = \mathbb{Z} * \mathbb{Z}$.

**Presenting a knot or link group.** (Wirtinger.) There is one generator for each arc of the knot projection. To each arc we associate a generator of $\pi_1(\mathbb{R}^3 - K)$ that makes a right-handed underpass for that arc. (This means as one moves along the generator, one can make a safe right-hand turn to get on the superhighway.)

Now write $ab$ for the loop that does $a$ first, then $b$. Then when $c$ crosses over $a - b$, so that we have a right-hand underpass, we get $ac = cb$. At a left-handed underpass, we get $ca = bc$.

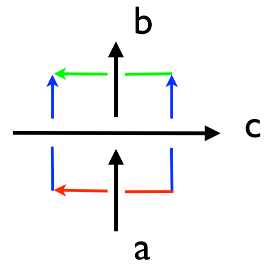



Figure 9. Proof that $ac = cb$ at safe underpass.

**Example: The trefoil knot group.** For a trefoil knot with all safe crossings, the group is

$$G(K) = \langle a, b, c \ : \ ab = bc, bc = ca, ca = ab \rangle.$$

**Mapping $G(K)$ to $\mathbb{Z}$.** This is easy: define $\phi(a) = \phi(b) = \phi(c) = 1$. It works for every knot group.

**The meridian.** Note that for a knot, *all generators are conjugate* in $\pi_1(K)$. This conjugacy class is given by the *meridian*, a small loop around the knot.

**Mapping $G(K)$ to $S_3$.** Notice that $S_3$ has 3 transpositions, call them $A, B, C$; they are all the odd permutations. They must satisfy $AB = CA$, since $ABA$ is odd, it can't be $A$ (else $A = B$) and it can't be $B$ (else $S_3$ is commutative). Mapping $(a, b, c)$ to $(A, B, C)$ we see $G$ admits $S_3$ as a quotient!

Corollary: The trefoil knot really *is* knotted.

**Tricolorings.** A tricoloring of a knot projection (without orientation) is the association of one of three colors (say $R, G, B$) to each strand, with the property that at each crossing either all 3 colors are the same or all 3 are different; and all colors are used.

**Theorem 15.6** *The tricolorings of a knot correspond one-to-one to the surjective homomorphisms $\phi : G(K) \to S_3$.*

**Proof.** Let $A, B, C$ be the three flips in $S_3$ as before. Since the generators of $G(K)$ correspond to arcs, we can use the three colors to define a map $\phi : G(K) \to S_3$ on the generators. Then the tricoloring condition shows that each relation in $G(K)$ is satisfied. So we can map the generators for strands of color $A$ to flip $a$, etc. Since at least two colors are used, we get a surjective map.

Similarly, if we have a surjection, then the generators (all being conjugate) must go to flips, else the image would be in the abelian subgroup of rotations. We then obtain a coloring. ■

Note: for a link the statement above is not quite true. A homomorphism that kills one component of a link does not correspond to a coloring. That is the tricolorings correspond to maps to $S_3$ that send all generators of the Wirtinger presentation to flips.

**Coda: links and carabiners.** The Hopf link $L$ is *less knotted* than the unlink, since $G(L) \cong \mathbb{Z} \oplus \mathbb{Z}$. As a trick, one can weave the commutator through two unlinked carabiner, in such a way that the loop comes free when the carabiner are linked!